Machine Autonomy and the Uncanny

Recasting Ethical, Legal, and Operational Implications of the Development of Autonomous Weapon Systems

Kjølv Egeland

Master's Thesis in Political Science

Department of Political Science

UNIVERSITY OF OSLO

Spring 2014

Machine Autonomy
and
the Uncanny

Recasting Ethical, Legal, and Operational Implications of the Development of Autonomous Weapon Systems

Machine Autonomy and the Uncanny

Kjølv Egeland

http://www.duo.uio.no

Print: Reprosentralen, University of Oslo

Illustration on the title page: Hephaestus, the Greek god of craftsmen. Wikimedia Commons, available at: http://commons.wikimedia.org/wiki/File:Hephaistos.png.

Abstract

While robots and *automata* have traditionally belonged to the realm of fiction, they are rapidly becoming an issue for the disarmament community. On the one hand, some experts believe that robots programmed to adhere to international humanitarian law (IHL) will be able to act more ethically than human beings on the battlefield. On the other hand, several commentators have disputed this claim, contending that the use of robots – or autonomous weapon systems (AWSs) – will lower the threshold to use violent force, and that such machines will be unable to discriminate between soldiers and civilians. Accordingly, this (essentially utilitarian) discussion of the consequences the deployment of AWSs is likely to have, remains locked in a word-against-word argument. Rather than focusing on the direct humanitarian effects of AWSs, people calling for a pre-emptive ban should point to the issue of moral agency, and the relationship between AWSs and human beings. *Machine Autonomy and the Uncanny* is an attempt at separating the question of 'harm' from questions pertaining to 'the harmer'. The use of AWSs poses grave problems for the doctrine of the moral equality of soldiers, for the dignity of all parties involved, and for both legal and moral responsibility.

Contents

Acronyms

4GW	Fourth generation warfare (guerrilla, terrorist, and immersion tactics)
API	Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977
APII	Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), 8 June 1977
AWS	Autonomous weapon system
CCW	Convention on Certain Conventional Weapons
DMZ	Demilitarized zone
HRL	(International) human rights law
HRW	Human Rights Watch
IAC	International armed conflict
ICC	International Criminal Court
ICJ	International Court of Justice
ICL	International criminal law
ICRC	International Committee of the Red Cross
IED	Improvised explosive device
IHL	International humanitarian law
LOAC	Law of armed conflict
NGO	Non-governmental organization
NIAC	Non-international armed conflict
RMA	Revolution in military affairs
UAV	Unmanned aerial vehicle (drone)
UN	United Nations

Preface

A Note on Terminology

'A rose by any other name would smell as sweet', Juliet argues in William Shakespeare's play.[1] Nonetheless, we need to pick *a* name for the objects discussed in this thesis. In the emerging literature on the subject, they are known by many: 'Military robot', 'killer robot', 'autonomous weapon system', 'lethal autonomous weapon', and their derivatives are all used to denominate the same objects, yet clearly hold different connotations. In my opinion, there are problems with all of these labels.

First, 'killer robot' seems to me best reserved for campaigners. A more objective term is to be preferred here. Second, 'robot' or 'military robot' I take to include a broader class of objects than the ones primarily under scrutiny in this thesis. Discarding the qualifying adjectives 'killer' and 'military' leaves us the terms in one way or the other including the word 'autonomous'. The term 'autonomous weapon system' is commonly used in the literature on which this thesis draws, but the phrase itself is seldom problematized.[2] In most moral philosophy, particularly following Immanuel Kant, autonomy would imply a form of free will, and moral and legal responsibility to go with it.[3] Yet, the prospects of robots at The Hague has not been seriously discussed by any of the authors using the term.

The Kantian definition of autonomy would advise that the term 'autonomous weapon' be left at the roadside.[4] However, it appears that the term is here to stay. Thus, I shall (reluctantly) employ it throughout this thesis. I do not, however, as some do, differentiate between *semi*-autonomous weapons and *fully* autonomous weapons. It seems to me that employing such

---

[1] W. Shakespeare, 2000, p. 94 (II.II).
[2] T. Hellström (2013, p. 101) and W. Wallach and C. Allen (2012, 2013, pp. 133–34) are notable exceptions.
[3] I. Kant, 2008, p. 57; L.A. Mulholland, 1990, p. 108–9.
[4] S.M. Shell, 2009, p. 2–3.

terms either renders the concept of 'automation' unnecessary, or becomes a tautology or pleonasm. The mixing of terms has contributed to the considerable confusion permeating the field.[5]

In this thesis, I take 'autonomy' to mean the capacity of a machine to act in a real-world environment independently of human control, including the capacity to make choices. 'Automated' or 'automatic' on the other hand, I take to mean the capacity to function according to pre-programmed mechanisms with reduced human interference. In this sense, a microwave oven that stops after a certain amount of time is automated, but not autonomous. While automation comes in degrees, autonomy is a *quality*.

Unless otherwise stated, I use 'autonomous weapon system' (AWS) and 'autonomous military robot' somewhat interchangeably. The meaning should be clear from the context. Neither do I enforce a sharp distinction between the terms 'ethics', 'morality', or any of their derivatives.

Acknowledgements

I am obliged to a number of people. With his enthusiastic guidance, my supervisor, Dr. Henrik Syse of the Peace Research Institute Oslo (PRIO), has been of great help. I also owe a large due to the co-supervisor, Dr. Gro Nystuen, the International Law and Policy Institute (ILPI), and the ILPI staff for academic and institutional support. As always, I have drawn on the succour of my brilliant parents, Marianne Egeland and Jon Haarberg. For the final result, however, I am alone responsible.

[5] See ICRC, 2013.

1 Introduction

'The robot soldier is coming',[7] *The New York Times* reported in 2005. For the first time ever, the darlings of science fiction appeared actual prospects for the conceivable future:

> 'They don't get hungry', said Gordon Johnson of the Joint Forces Command at the Pentagon. 'They're not afraid. They don't forget their orders. They don't care if the guy next to them has just been shot. Will they do a better job than humans? Yes'.[8]

The last few decades have seen an explosion in use, research, and development of increasingly automated weapon systems. The development is driven both by government agencies, such as the American Office of Naval Research (ONR) and Defense Advanced Research Project Agency (DARPA), and by private companies across the world, like the South Korean Samsung, the American Northrop Grumman, and the British QinetiQ.

Remote-controlled armed drones being well underway on their third decade in use,[9] the advent of *autonomous* weapon systems appears to be drawing near.[10] Such systems – not yet in existence – would be capable of identifying, selecting, and engaging targets without direct human interference. It is with such weapon systems that this thesis is concerned.

Information and communication technologies have become crucial components in coordinating military operations. However, using such technologies exposes a modern army to cyber-attacks and other attempts at inhibiting information-flows, communication, and control. In this context,

---

[6] Latin proverb: 'War is a treat for those who have not tried it.' See Erasmus, 1982, pp. 399–400 (*Adages* IV i 1).
[7] T. Weiner, 2005.
[8] T. Weiner, 2005.
[9] The first use of armed drones is believed to have been carried out by Iran against Iraqi ground forces in the late 1980s. F. Haghshenass, 2008, p. 17. The United States military has researched drones since 1917. J. Gertler, 2012, p. 1.
[10] R.C. Arkin, 2008.

autonomous weapons have a crucial role, being envisioned as less susceptible to hacking, blocking, and spoofing:[11]

> More sophisticated […] opponents could adopt counter strategies that would render drones useless by jamming communication signals. But a fully autonomous drone could seek out its target without having to communicate with an operator. The use of autonomous drones is also likely to reduce military costs and the number of personnel required, and to improve operation by stripping out human error and responsetime limitations.[12]

Autonomous systems would moreover be better suited for air battle. Due to the delay associated with processing information through satellites and control centres, current drones have a time lag, or 'latency', between the operator pulling the trigger, and the drone firing its missile. AWSs, on the other hand, would not suffer from this weakness.[13]

A new US Military Concept called 'Air–Sea Battle' is seemingly a plan for a large-scale international armed conflict against a high-tech sate, probably in and around the Pacific and China Seas.[14] In such a hypothetical conflict, autonomous weaponry could be of vital importance. As the most recent US roadmap for unmanned military technologies acknowledges, a potential conflict in the Asia–Pacific Theatre presents new challenges, potentially requiring unmanned systems to operate in 'anti-access/area denial (A2/AD) areas where freedom to operate is contested'.[15] Funding for unmanned systems by the US Department of Defense was increased from 284 million USD in 2000 to 4.2 billion USD ten years later.[16]

AWSs have quickly become an issue in international forums, such as the United Nations Human Rights Council,[17] the United Nations General

---

[11] Allegedly, a US drone was hacked and brought to the ground by the Iranian army in 2011. See J. Mick, 2011.
[12] N. Sharkey, 2013.
[13] K. Abney, 2013, p. 343.
[14] Air–Sea Battle Office, 2013, p. 4. See T. Cobb, 2011.
[15] US Department of Defense, Unmanned Systems Integrated Roadmap, 2013, p. v.
[16] US Department of Defense, Unmanned Systems Integrated Roadmap, 2013, p. 3.
[17] United Nations Office for Disarmament Affairs, 2013.

Assembly First Committee,[18] the European Parliament,[19] and, perhaps most notably, in the framework of the Convention on Certain Conventional Weapons (CCW).[20] The CCW expert discussions were attended by a large number of states parties, including the all the permanent members of the United Nations Security Council.[21]

With a vote of 534 to 49 on 25 February 2014,[22] the European Parliament adopted a resolution calling on the EU's member states, the High Representative for Foreign Affairs and Security Policy, and the European Council to 'ban the development, production and use of fully autonomous weapons'.[23] Given the relative weakness of the EU Parliament compared to other organs, the Resolution was passed without much attention. Nevertheless, the Resolution testifies to the gained traction of AWSs on the international agenda. An international campaign against AWSs, dubbed 'Stop Killer Robots', was launched by a large group of NGOs in 2012.[24] They claim that the use of AWSs is unethical, and should be prohibited.[25]

Countering the critics of AWSs, the Georgia Tech roboticist Ronald C. Arkin has argued that properly programmed robots could in fact be better than human beings at adhering to the laws of war. Robots, he explains, do not feel pain, anger or lust of revenge, possibly the most common causes of war crimes. Moreover, removing human beings from the battlefield (at least on one side of the equation), would, *ceteris paribus*, lead to a net reduction of physical and mental trauma, he argues.[26]

The US Department of Defence issued a directive in 2012, determining that all US strikes should be controlled and confirmed by human operators. While the Directive appeased some critics, the promise is due to expire in

---

[18] United Nations General Assembly First Committee official website, 2013.
[19] European Parliament Resolution on the Use of Armed Drones, 2014.
[20] United Nations Office for Disarmament Affairs, 2013.
[21] M. Brehm, 2014, p. 2.
[22] R. Goodman, 2014.
[23] European Parliament Resolution on the Use of Armed Drones, 2014, Paragraph 2(d).
[24] See S. Goose, 2013.
[25] S. Goose, 2013.
[26] R.C. Arkin, 2008, pp. 6–8. See also M. Sassòli, 2013.

2022.[27] Furthermore, the promise to stay 'in the loop' appears contradictory to the conclusions of the Roadmap for Unmanned Weapon Systems[28] and (if more implicitly) the Air–Sea Battle Concept,[29] both of which stress that autonomy may be militarily necessary.

Most successful campaigns to ban specific weapons, like the campaigns against anti-personnel landmines and cluster munitions, were able to build their cases on the basis of grave and well-documented violations of the rule of discrimination between soldiers and civilians.[30] However, it is unclear whether such arguments can be made for AWSs. As they are not yet in existence, the actual consequences of their use are difficult to assess. It has been argued that the Campaign to Stop Killer Robots lacks a central argument: What is actually at stake? What is the case against 'killer robots'?[31] According to one commentator, the case presented by Human Rights Watch (HRW), founding and senior member of the campaign Stop Killer Robots, is dystopian and irrational:

> By reducing the discussion to a simplified fear campaign aimed at stopping the development of autonomous systems, HRW may very well be preventing the deployment of systems that can be more careful and less likely to harm civilians. Such an approach means their fear campaign may end up harming the innocent people they intend to protect.[32]

In this thesis, I investigate and expand on the case against AWSs. Thematically, the topic of this thesis touches upon the disciplines of ethics, law, and political science. Accordingly, my methodological approach is eclectic (see Chapter 2). All of the sub-questions addressed in this thesis could arguably be dealt with more thoroughly. In this thesis, however, the idea is to

---

[27] US Department of Defense, 2012, p. 4.
[28] US Department of Defense, Unmanned Systems Integrated Roadmap, 2013, p. v.
[29] Air–Sea Battle Office, 2013, p. 2.
[30] It has been argued that the apparent lack of clear military utility of the weapons in question was also important. E.g. ICRC, 1994, p. 73; O. Dullum, 2007, p. 143. A notable exception to the notion that banning a weapon requires well-documented violations of IHL and HRL is blinding lasers, which were banned pre-emptively through the CCW Protocol IV.
[31] E.g. G. McNeal, 2013; S. Hughes, 2013.
[32] G. McNeal, 2013.

present a policy-oriented birds-eye perspective, while still offering new approaches to the development and use of AWSs.

1.1 Research Question: How Might a Prohibition of AWSs Be Justified?

The heading above captures the central research question of this thesis: *How might a prohibition of autonomous weapon systems (AWSs) be justified?* Answering this question will require an identification and assessment of normative arguments applicable under existing international law. Until now, all conventions banning particular weapons have been justified with the weapons' breach with the norms of distinction or of avoiding unnecessary suffering and superfluous injury.[33] If current international law and its history is anything to go by, this thesis could be limited to addressing those two norms. However, AWSs raise concerns that until recently were inconceivable. What a legal ban really comes down to is political will. Thus, I believe it both important and necessary to look at arguments beyond the traditional requirements of distinction and avoidance of unnecessary suffering and superfluous injury. I shall argue that while international law is an essential component in our evaluation of these weapon systems, existing law fails to address the core challenge posed by AWSs.

In order to answer the overarching research question, four main questions have to be addressed along the way:

[33] To the former category belong weapon (systems) such as anti-personnel mines, and cluster munitions. To the latter category belong weapon (systems) such as incendiary weapons, blinding lasers (the only weapon to be banned pre-emptively, i.e. before it was ever in use), and non-detectable fragments. Some weapons could be argued to be banned on both grounds, for example chemical and bacteriological weapons.

Part I: 'Harm'

(1) A slippery slope?: Do unintended, negative consequences of the use of AWSs justify their prohibition?
(2) The conduct of hostilities: Can the use of AWSs be squared with the rules of distinction, proportionality, avoidance of unnecessary suffering, and precautions in attack?

Part II: 'The Harmer'

(1) Accountability: Who is responsible for potential international crimes committed by an AWS?
(2) Agency: Is it morally permissible to delegate kill-decisions to machines regardless of the humanitarian consequences of their use?

As is evident from the classification above, I propose to divide the thesis in two, based on the perspective taken. In the fist part, I address the humanitarian consequences of the use of AWSs. In the second part, I leave consequences aside, discussing the responsibility and agency of autonomous systems.

1.2 Definition and Background

There is no universally agreed upon definition of AWSs. The discussion in this thesis is based on the ICRC's working definition, which, although terse, seems to capture the central element that states and academics discuss when referring to AWSs. They define AWSs as:

> *Weapons that can independently select and attack targets.*[34]

The word 'select' is a key word of the definition, as it implies the capability to make intelligent decisions. A 'target' should be understood as an enemy soldier or military object the sense of international humanitarian law (IHL); 'civilian target' is a tautology.[35] Selecting a target necessitates sophisticated information processing and judgement. In the language of computer science and robotics, an autonomous agent requires significant 'artificial intelligence' or

[34] ICRC, 2014b, p. 1. The ICRC's definition is very similar to those of Human Rights Watch (2012, p. 2), Pax for Peace (2014, p. 4), and the US Department of Defense (US DoD Directive, 2012, p. 14).
[35] API, Article 51. See S. Casey-Malen, 2014 (forthcoming).

'autonomous power'.[36] Hence, the definition excludes for example anti-personnel landmines. Crucially, an AWS is capable of selecting targets 'independently', *i.e.* without human intervention. I shall not discuss the use of military robots only intended for such tasks as transportation, mine detection, or bomb disposal. Although some of the inferences made in this thesis may apply to any autonomous system, I am primarily concerned with weapons capable of harming human beings. I am less concerned with weapons only intended to engage inanimate objects, for example missile defence systems.[37]

Noel Sharkey provides a classification of three types of automated systems, the latter of which may be considered autonomous according to the definition above: (1) *scripted*, like intercontinental ballistic missiles, (2) *supervised*, like RQ-1 Predator drones, and (3) *intelligent*, 'which uses attributes of human intelligence in software to make decisions, perceive and interpret the meaning of sensed information, diagnose problems, and collaborate with other systems'.[38]

In the literature on AWSs, reference is commonly made to the concepts of having humans '*in* the loop', '*on* the loop', or '*out of* the loop'.[39] While 'in the loop' implies that the weapon system is remote-controlled by a human, 'on the loop' means that a human does not directly control the weapon system, but surveys its actions, and can assume direct control at any time. Lastly, 'out of the loop' Is taken to mean an arrangement whereby the weapon system acts autonomously, without any human interference. In the terminology of this thesis, AWS is synonymous with the latter.

It is generally agreed that AWSs do not yet exist, but there are clear precursors.[40] According to most definitions, AWSs are able to 'attack targets'. While no technology is failsafe, the ability to attack a 'target' implies at a minimum that the weapon is capable of distinguishing between combatants and

---

[36] T. Hellström, 2013, p. 101.
[37] E.g. the US Navy's Phalanx CIWS.
[38] N. Sharkey, 2008, p. 16.
[39] E.g. Sharkey, 2010, p. 369; Human Rights Watch, 2012, p. 2; M.N. Schmitt and J.S. Thurnher, 2013.
[40] See C. Heyns, 2013, p. 5; Human Rights Watch, 2012, p. 3.

non-combatants on a rudimentary level: No non-combatant may be described as a 'target'. For example, Samsung's SGR-A1 and DoDAAM's Super aEgis 2, both stationary sentry gun systems stationed at the Korean 'demilitarized' zone, are able to identify and lock onto targets without human operators.

Raytheon's stationary Phalanx CIWS (close-in weapon system)[41] is according to its manufacturer capable of carrying out 'functions usually performed by multiple systems – including search, detection, threat evaluation, tracking, engagement, and kill assessment'.[42]

A new generation drone, Northrop Grumman's X-47B, is currently undergoing flight tests, and is expected to be operational in 2015. In the fall of 2013, it became the first vehicle to successfully complete a flight entirely by algorithm (without human controllers), including landing and take-off.[43] Whether the system also has the capacity to select and engage targets autonomously, has not been disclosed.[44]

Yet, neither the SGR-A1, Phalanx CIWS, X-47B, or any other weapon system have – according to most accounts[45] – been used to full effect with humans 'out of the loop'. They are furthermore utterly incapable of distinguishing between military targets and civilians, thus falling short of the definition.[46] On the other hand, there is obviously a large grey area. For example, with some automated systems, the time allotted for human deliberation 'in the loop' – the decision whether to allow the system to fire – is sometimes, as with most robotic anti-missile defence systems, a matter of a few seconds, and is thus not very meaningful.[47] Moreover, with 'signature' drone strikes, targeting is based on predetermined criteria, and the role of human

[41] The Phalanx system was developed by General Dynamics Corporation which was later bought by Raytheon. It 'is installed on all US Navy surface combatant ship classes and on those of 24 allied nations'. See Raytheon's official website, 2014.
[42] Raytheon's official website, 2014.
[43] Naval Air Systems Command, 2013.
[44] D. Gayle, 2012.
[45] See M.N. Schmitt and J.S. Thurnher, 2013, p. 239.
[46] See M.N. Schmitt and J.S. Thurnher, 2013, p. 239.
[47] P.W. Singer, 2010, pp. 124–5.

intelligence and decision-making is severely limited.[48] Evidently, we see that while the distinction between autonomous and automated weapons is unproblematic in theory, in practice it becomes blurred. Nevertheless, discussing AWSs – the 'ideal type' of a machine having control of targeting-decisions – is important in order to identify the principles and ideals against which we would like to hold both existing and emerging technologies.

Both the academic and the public discourses have been hampered by imprecise use of terms and concepts, and by a failure to be explicit about the types of weapons and platforms actually under scrutiny. While there are common issues to both unmanned, automated, autonomous, or even the broader range of emerging military technologies as a whole, autonomous weapons beg a host of additional, distinct questions of their own, particularly regarding target selection.[49]

Over the last decade, military robots have received heightened interest. The obvious reason for this is the simple fact that due to amazing advances in electronics and engineering, science fiction is no longer as fictitious as it used to be. 'Revolutions in military affairs' (RMAs) are well documented through history, signifying the introduction of ideas or equipment into the field of war with consequences so profound that 'the rules of the game' have to be re-written. Has the development in software engineering and robotics now produced a robotic RMA?[50]

So-called riskless warfare is not new. Catapults and artillery are old inventions that, at least under asymmetric conditions, have allowed the risk-free inflection of harm. However, AWSs differ from former long-distance weapons in important ways: While some robotic systems are intended to *complement* human soldiers, others are intended as their *replacements*. Such a role is qualitatively different from that of former 'riskless' weapons, which

---

[48] Stanford and NYU, 2012, p. 12.
[49] For example, the issue of sovereignty and non-intervention is commonly raised in the context of armed drones. Since this issue does not appear to have direct relevance to autonomous targeting, it falls beyond the scope of this thesis.
[50] E.g. P. Rose, 2012, p. 366.

more strictly are tools in the hands of humans. In the case of these older weapons, there is a clearer causal link between humans and the firing of the weapon. In the case of AWSs, the system itself both selects and attacks its target, evidently breaking the direct causal link between human and weapon.

2 A Poverty of Normative Methods?

This thesis belongs in part to the field of ethics and in part to the field of jurisprudence or legal philosophy, both of which may be said to belong to the broader field of normative political philosophy. Analysing the case against AWSs call for the use of normative and legal research methods. Any study is arguably conducted by means of some method, but are there distinctive methods for normative theory? In *Politics and Vision*, Princeton professor Sheldon Wolin writes that political philosophy distinguishes itself from other branches of academia mainly through its subject matter. The philosophical method and its goal is the same for all its schools and branches: 'truths publicly arrived at and publicly demonstrable'.[51] In practice, this implies logical reasoning and convincing arguments. At the same time, this is admittedly the foundation of every scholarly discipline there is. In an influential article, Wolin went on to identify a conflict between 'methodists' and 'theoreticians'. He asserted that there is an opposition between method and theory, and that students of politics have to pick one of the two[52] I would argue that there is no such built-in opposition between the theory and method. A method is simply the way in which one generates, corroborates, or tries to falsify theory, or, more plainly, the way in which one practically goes about answering questions. In this section, I first discuss the philosophical methods of the dialectic, interpretation, and non-ideal assumptions, before I briefly present the legal positivist method, which informs the legal parts of this thesis.

2.1 Dialectic

First, I regard the principle of contradiction advocated by such thinkers as Socrates and Hegel as a favourable way of approaching the ideal of objectivity. A schematic approach of 'thesis–antithesis–synthesis', however, is not helpful,

[51] S.S. Wolin 2004, p. 4.
[52] S.S. Wolin, 1969, p. 1062.

and was advocated neither by Plato's Socrates nor Hegel.[53] Hegelian dialectics is wedded to the thought that the contradiction of a concept is a fruitful way of broadening the analyst's mental space (and somewhat more controversially how the external world is structured).[54] In the terminology of political science we might say that the dialectical method is a philosophical way of ensuring the largest possible variation on the variables in order to avoid bias.[55]

According to Christopher Meckstroth, the whole concept of a 'method', and indeed the Greek word *methodos*, first appears in Plato's writings. He sums up Socrates' principles as '(1) anti-foundational, (2) non-algorithmic, (3) indirect and relative to competing hypotheses, and (4) developing its own standards of objectivity through the logic of asking questions'.[56] Plato calls the method *elenchus*, which could be translated as meaning 'cross-examination' or 'refutation'. The posing of questions force us to think twice about the coherence of our beliefs, and how principles relate to each other.

Socrates arguably does not start from some first principle or theory of 'the good', he simply forces his ever-confident dialogue partners to think about how their principles relate to each other, and to the empirical world. The method is thus based on negatively identifying internal inconsistencies, often labelled immanent critique. This technique distinguishes itself from both deduction and induction as it does not presuppose any positive foundational premise, be it logic or statistical inference.

The idea that sound ethical injunctions can be elicited through a dialogue between people taking opposite positions can be extended to the relationship between an analyst and his or her textual sources.[57] This is how the dialectical method is most often employed in practice, and is how I shall practice it in this thesis. Ethical propositions and statements made in books and

---

[53] E.g. G.E. Mueller, 1958.
[54] C. Taylor, 2005, p. 226; G.W.F. Hegel, 1963I, pp. 30–1.
[55] Bias resulting from truncated variables arises when cases are selected on the dependent variable; that is, when only cases with either high or low scores on the dependent variable are included in the analysis. A truncated independent variable on the other hand does not cause bias, but 'only' imprecision (increased standard errors). See B. Geddes, 1990.
[56] C. Meckstroth, 2012, pp. 644–5.
[57] H.G. Gadamer, 1999, pp. 362, 365, 368.

articles can be tested and interpreted by demonstrating their coherence and relationship to broader principles or codes. However, any one 'truth' can never be obtained using the dialectical method; what one is left with are reasons for preferring one explanation or statement to others. As such, the output of the dialectical method is 'true' or 'false' only *relative* to competing positions.

Just as empirical science, normative scholarship is grounded in empirical facts and theoretical directives. Normative theorists make claims about how the world should be ordered and why, starting in Ancient Greece with how the just city should be ordered.[58] Thus, they are just as bound to aspire to valid and reliable knowledge about how the world works as analysts with more strictly empirical pretentions. As David Hume so famously argued, 'is' does not imply 'ought', but 'ought' to a certain extent implies 'can'.[59] The 'ought'-part of normative theory, however, (indeed its defining part) will be chiefly theoretical (although the theory builds on empirical facts). How to arrive at the 'ought' will, methodically, to a large degree be determined by the theory to which the author subscribes.[60] Hence, in ethics, the overlap between theory and method is perhaps greater than in more strictly empirically oriented disciplines.[61]

An epistemological distinction manifest in modern political philosophy is also worth mentioning: Theorists such as John Rawls, Michael Walzer, and Axel Honneth aim to approximate objectivity by coming up with justifications for arguments through analytic or dialectical arguments. The members of another group, to which Judith Butler, Richard Ashley and others may be said to belong, hold that analytical arguments conceal important questions about politics and power, rather preferring deconstructive, post-structural, and other methods agonistic to objectivity.[62] I take the position of the former – that it is indeed both possible and desirable to make sense of the world with the

---

[58] Plato, 1989, see especially books IV and V.
[59] See G.E.M. Anscombe, 2007, pp. 30–1.
[60] See N.L. Sturgeon, 2011, pp. 517–9.
[61] See D. Leopold, 2010, p. 126.
[62] C. Meckstroth, 2012, p. 644.

language available to us in an attempt at approximating objectivity. This is not to say that we should not be critical of the terms we use.

2.2 Interpretation

In his *Interpretation and Social Criticism*, Michael Walzer differentiates between three ways of doing moral philosophy: discovery, invention, and interpretation. The path of *discovery* is often employed by theorists of natural law and by many religious thinkers. Here, ethical propositions are thought to be revealed or discovered. They are in other words, created not by men or women, but by the Divine or by inspiration.[63] Using the path of *invention*, the analyst or inventor sets out to detach herself as much as possible, entering the position of the eagle (or, by analogy, the position of the Divine). For the output of such a method to represent more than the subjective injunctions of an individual, the inventor must assume an impartial position. The problem of how one might achieve this in practice was most elegantly solved by John Rawls with his original position, the 'veil of ignorance',[64] Walzer asserts.[65]

The third and final path, to which both Walzer himself and this author subscribe, is the path of *interpretation*. 'The claim of interpretation is simply this', Walzer asserts, 'that neither discovery nor invention is necessary because we already possess what they pretend to provide. Morality, unlike politics, does not require executive authority or systematic legislation'. Indeed, 'the moral world has a lived-in quality'.[66] Normative theory should start from the values manifest in society, rather than formal, abstract principles.[67] I take the same methodological approach, assuming that the application of abstract principles to society that the inventor or discoverer must make, often complicates questions of methodology and viability more than it brings in clarity and

[63] M. Walzer, 1987, p. 4.
[64] See J. Rawls 1999.
[65] M. Walzer, 1987, pp. 4, 11.
[66] M. Walzer, 1987, pp. 19–20.
[67] A. Honneth, 2011, p.119.

objectivity. Yet, as with the distinction between ideal and non-ideal theory discussed in the next section, the distinction between the three paths is by no means watertight.

The distinction between the path of interpretation and the paths of discovery and invention to a certain extent mirrors a corresponding debate of about two hundred years ago. In his 1821 *Philosophy of Right*, G.W.F. Hegel accuses Immanuel Kant of 'empty formalism', pointing to the practical infeasibility of the abstract categorical imperative. The imperative – 'act only according to that maxim, whereby you could at the same time want that maxim to become a universal law'[68] – is only as good as the person contemplating it, Hegel argues.[69] While recognizing that the radical autonomy of the categorical imperative could serve analytical purposes, Hegel held the actual norms of a society, its 'ethical idea' (*Sittlichkeit*), to be the pinnacle of social legitimacy, authority and obligation.[70]

There is at least one important objection to Walzer's and Hegel's stress on the here-and-now as the ultimate appeal of ethics: It might neglect a significant *critical* aspect of ethical inquiry.[71] Looking back at history, it is obvious that the moral fibre of any society evolves; we are fervently opposed to many practices formerly endorsed or unquestioned. Obvious historical examples include slavery and suppression of women and many minorities. One of the founders of critical theory, Max Horkheimer, warns against a theory that 'settles down with some truth'[72], and advises the analyst to reflect critically on the relationship between the subjective and the objective, between the analyst and her objects, and between concrete historical processes and the future.[73]

Indeed, the prospect of moral relativism lurks in the background of any communitarian or interpretative theory. In response to this, I would second the

---

[68] My translation of 'Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daβ sie ein allgemeines Gesetz werde', I. Kant, 2008, p. 57.
[69] G.W.F. Hegel, 1902, §§ 140-1.
[70] G.W.F. Hegel, 1902, §§, 70, 257.
[71] A. Honneth, 2003, p. 137.
[72] My translation of 'Die Philosophie, die bei sich selbst, bei irgendeiner Warheit, Ruhe finden meint'. M. Horkheimer, 1970, p. 64.
[73] M. Horkheimer, 1970, p. 24.

theorists adhering to the so-called theory of recognition in their claim that – albeit with constant setbacks and inconsistencies – the moral evolution of societies by way of intersubjective recognition is a positive one. It should, then, be possible to unveil immoral practices or inconsistent positions if one is able to anticipate the direction of this evolution.[74]

2.3 Non-Ideal Assumptions

An important distinction that has received increasing attention in the literature on the methodology of political theory is between so called ideal and non-ideal theory. The distinction was first made by John Rawls,[75] and concerns the rudimentary assumptions the analyst makes when investigating an issue. Whereas ideal theory deals with the ancient question of how to build a just city from a *tabula rasa*, non-ideal theory is concerned with how to approximate justice in practice.[76] According to critics, proponents of ideal theory do not merely abstract away complicating factors, but add unrealistic assumptions, thus being guilty of 'idealization'.[77] Non-ideal theory, on the other hand, proceeds from the assumption that the real world tends to require pragmatic solutions.[78]

The proponents of ideal theory are not defenceless against this criticism. In order to imagine the perfect system, or justice in its ideal form, we might have to assume what Michael Walzer labels 'the position of the eagle', looking down at the here-and-now from a privileged position.[79] If several analysts make the same assumptions, moreover, it is possible to build cumulative scholarship also in normative disciplines.[80] According to Zofia Stemplowska, the relationship between ideal and non-ideal theory is not dichotomous. Rather, it

---

[74] A. Honneth, 1995, 2003; C. Taylor, 1994. See Chapter 6 for more on the theory of recognition, and its application to autonomous weapon systems.
[75] J. Rawls, 1999, pp. 215-6.
[76] A. Swift and S. White, 2010, p. 58.
[77] C. Farrelly, 2007, pp. 844–5.
[78] C. Farrelly, 2007, pp. 844.
[79] M. Walzer, 1987, pp. 4, 11.
[80] See Z. Stemplowska, 2008, pp. 327–8.

is a continuum between making a large amount of fanciful abstractions and making none at all. All points along this continuum may have their strengths and weaknesses, and most analysts, for instance Rawls, make use of both ideal and non-ideal theory for different sections of their work.[81]

Taking on questions about the use of autonomous weapon systems in international armed conflict, I have to consider the broader system in which these questions are to be placed. This broader system, or context, is international relations. According one of the most dominant theoretical conceptions of this system, the school of realism, the international system of states comprises an anarchy. The lack of a world government causes the states to fear each other's intentions, producing the behavioural outcome of self-help and a desire to maximize one's power.[82] Since there is no world government, states are left to their own best efforts.[83] Although this Realist analysis has the air of a pointed formulation, there appears to be at least some truth to it. This has led many to question the vigour of international law and cooperation in international affairs. States are unlikely to bind themselves or stay bound if the mechanism in question is not conducive to their national interests (in which humanitarian concerns feature only to varying degrees).

In this context, it seems obvious that the international system is not ideal. Using ideal theory to analyse robotic warfare, then, a sub-issue within international relations, would be an academic exercise bordering irrelevance, if done without constructing an ideal theory for the entire international system. Such an exercise falls beyond the scope of this thesis. I would argue that both ideal and non-ideal theory have their virtues. For the purpose of this thesis, however, non-ideal assumptions seem to me more appropriate. Yet it must be stressed: I do not think that it is possible to write without – implicitly or explicitly – making assumptions. Assumptions are never 'true', they are

---

[81] Z. Stemplowska, 2008, p. 339. See also A. Hamlin and Z. Stemplowska, 2012, p. 60.
[82] A distinction is sometimes drawn between offensive (favouring power maimation) and defensive realism (favouring security maximation), but this is not important in this context. See e.g. G.H Snyder, 2002, p. 149.
[83] J.J. Mearsheimer, 2001, pp. 33–4.

heuristic devices to help us make sense of the complex and messy world in which we live. As the Hungarian philosopher of science, Imre Lakatos, puts it: Without assumptions and theory, scholarship becomes 'a curiosity shop where funny local – or cosmic – oddities are collected and displayed'.[84]

The theoretical lenses employed in this thesis are Hegelian. All but totally absent from the largely Anglo-American peace and conflict literature and the literature on weapons and disarmament, Hegel has been revived in social theory and philosophy. The current director of the Institut für Sozialforschung at the University of Frankfurt, Axel Honneth, has been pivotal in this rediscovery and development of the dialectical Hegelian thought in modern social theory. There may well be a reason for the absence of Hegelian thought in the recent peace and conflict literature. According to Karl Popper, Hegel's thought is no more than 'emptiness': 'The question arises whether Hegel deceived himself, hypnotized by his own inspiring jargon, or whether he boldly set out to deceive and bewitch others. I am satisfied that the latter was the case'.[85] Notwithstanding this critique, I shall try to make sense of Hegelian thought and its application to AWSs, particularly in the last chapter on the question of AWSs with a 'licence to kill'.

Normative statements are always opined. Nonetheless, political philosophy has a strong methodological basis, and there are indeed normative methods that go beyond 'mere' argument and reason. In this thesis I work primarily from the basis of non-ideal assumptions, and I shall try to interpret the moral arguments put forth in the literature in a critical, dialectic fashion. The theoretical frameworks employed in this thesis are not presented in one bulk, but is drawn upon during the course of the analysis.

[84] I. Lakatos, 1984, p. 102.
[85] K. Popper, 1945, p. 26.

2.4 Legal Positivism

The parts of this thesis in which the legality of AWSs under existing international law is discussed (chapters 4 and 5), I draw on so-called legal positivism, the dominant methodology when determining *lex lata*, or 'the law as it is'. Analysts sometimes also discuss *lex ferenda*, or 'how the law should be'. Notwithstanding the political and moral importance of discussing *lex ferenda*, courts and legal practitioners are bound by *lex lata* when delivering judgements. Thus, when determining the existing landscape of the law applicable to autonomous systems, I employ legal positivism. The other parts of this essay, in which potential grounds of banning AWSs beyond international law is investigated, may reasonably be labelled a study of *lex ferenda*.

Legal positivism has its roots in the philosophy of Jeremy Bentham and John Austin, but received its most famous formulation by H.L.A. Hart. Hart primarily discussed domestic law, in which the structure of enforcement is typically envisioned as vertical, with a 'Leviathan'[86] on top of the structure. International law on the other hand, should be thought of as a horizontal system, founded on agreements between consenting states. However, it is generally agreed that most of Hart's conclusion are applicable also to international law despite the difference in enforcement structure.[87]

According to John Gardner, legal positivism can be summarized as the following proposition: 'In any legal system, whether a given norm is legally valid, and hence whether it forms part of the law of that system, depends on its sources, not its merits'.[88] Accordingly, I shall draw on the law itself, as custom and as codified in treaties and conventions, and on the comments of legal scholars and Courts. The Statute of the International Court of Justice defines the following as the sources of international law:

---

[86] See T. Hobbes, 2008, Chapter XIII.
[87] H.L.A. Hart, 1961, pp. 14–5, 217–20. See also S. Besson and J. Tasioulas, 2010, p. 9–11.
[88] J. Gardner, 2001, p. 199.

> (a) international conventions […]; (b) international custom […]; the general principles of law […]; (d) […] judicial decisions and the teachings of the most highly qualified publicists of the various nations, as subsidiary means for the determination of rules of law.[89]

These shall be the primary sources for my discussion on the legality of AWSs under existing international law. It is important to note, however, that legal positivism informs this study as a method of determining *lex lata*, not as a theory to judge right from wrong or determine ontological or epistemological questions beyond those of existing international law.

[89] Statute for the International Court of Justice, Article 38.

PART I: HARM

3 A Slippery Slope?

Let us assume, for the purpose of this chapter, that robots could function in the manner described by Ron Arkin, who is convinced that AWSs could perform 'more ethically than human soldiers'.[90] (This question is dealt with more thoroughly in Chapter 4.) There are nevertheless objections to the use and development of AWSs. In this chapter I investigate a class of arguments stressing that the deployment of AWSs may lead to a host of unintended consequences. The aim here is not to exhaust the issue of unintended consequences, but to hint at a few challenges a Consequentialist 'slippery slope' approach faces.

Breaking an existing rule, or introducing a new one, might in some cases lead to the watering down or disregard for other rules, or have unintended, negative consequences. This is commonly known as the 'slippery slope' argument. The technological introduction of AWSs does not exactly introduce a legal rule, or necessarily conflict with old ones (see Chapter 4), but as a potential RMA, it constitutes a change in 'the rules of the game', and impacts on how we think about war.

3.1 A Lower Threshold to Use Force?

The 'slippery slope' argument is commonly rehearsed against the use and development of AWSs. Peter Asaro, for example, claims that AWSs have the potential to 'lower the threshold for nations to start wars'.[91] In addition to impacting the initiation of war, the use of AWSs may impact the threshold to resort to the use of force in already on-going conflicts. While the former is a

[90] R.C. Arkin, 2008, p. 7; M. Sassòli, 2013. Being better at adhering to international humanitarian law, however, does not necessarily imply a net decrease in overall harm. Killing enemy soldiers is perfectly in line with IHL (API, Article 48), and the proportionality principle allows for unlimited civilian casualties as long as they can be justified in military advantage (API, Article 51).
[91] P. Asaro, 2012, p. 692.

question of *jus ad bellum* (the rules for the initiation of war), the second is a question of *jus in bello* (rules for the conduct of war).

In this section, I draw on the just war tradition, a long line of thought dating back to ancient Greek philosophy.[92] The just war tradition may be identified as one of three very broad takes on war: First, the programmatic Realists view is that 'universal moral principles cannot be applied to the actions of state'.[93] A classic exposure of realist reasoning is found in Thucydides' *History of the Peloponnesian War*, where the conquering Athenians explain the Melians that 'the strong do what they can and the weak suffer what they must'.[94]

Second, antithetical to the realist stance, pacifists oppose war more or less categorically. Many pacifists have for example opposed the development of international humanitarian law (IHL), because it implicitly recognized the act of using force.[95]

While pacifism and realism (somewhat simplified) are absolutes on each end of a spectrum, thinkers employing a third approach, the just war tradition, have typically claimed that wars can be legitimate granted that certain conditions are met, including a just cause, a right authority, and proportionality. According to Oxford professor Henry Shue, the attitude of IHL towards the grave violations of ordinary law and morality that war, killing, and wounding imply, 'can be well captured by the pithy phrase "shit happens"'.[96]

Augustine of Hippo is an important figure in the just war tradition. His attempt to bridge the messages of the New and Old Testaments resulted in the view that while Christians should on principle be pacifists, they were

---

[92] E. Begby *et al*., 2012, p. 316.
[93] H. Morgenthau, 2006, pp. 9–10.
[94] Thucydides, 2009, V.XVII. Emphasis added.
[95] R.P. Alford, 2008, p. 68. Strict Realist and Pacifist positions typically obligate their subscribers to strong non-ideal and ideal methodological positions respectively. IHL – which largely grew out of the just war tradition of Augustine of Hippo, Thomas Aquinas, Hugo Grotius, and others – lends non-ideal assumptions from Realism, and the moral content – humanitarianism – from Pacifism. See G. Nystuen and K. Egeland, 2014.
[96] H. Shue, 2010, p. 516.

nevertheless justified in taking up arms if they needed to protect themselves or their religion from unruly neighbours.

3.1.1 The Limitations on the Use of Force *ad Bellum*

Due to the harm it causes, Michael Walzer argues, war should always be a last resort.[97] The argument is one of the central tenets of the just war tradition.[98] According to several of the most quoted authors in the field, including Armin Krishnan, Peter Singer, Human Rights Watch, Pax, and Peter Asaro, the use of AWSs may create lower barriers for war through the technological benefits they present.[99] While none of these authors discuss the proposition in any detail, a lower threshold to use force, we must assume, is an ethical challenge because conflict inevitably leads to the infliction of harm. Two questions are of importance: First, is it true that AWSs lower the threshold to use force? And second, if the first question is answered affirmatively, does this mandate the prohibition of AWSs? Let us first consider the mechanisms that limit the freedom of state leaders to use force (*i.e.* heighten the threshold to start wars).

First, international law prohibits all use of armed force unless in self-defence or if authorized by the UN Security Council.[100] State leaders pondering the use of armed force would have to consider legal ramifications and, as a future prospect, potential criminal prosecution of their acts.[101] Second, state leaders have beliefs, values, and moral understandings, which may or may not restrain their will to wage war. Third, wars are financially expensive and cost

---

[97] M. Walzer, 2006, p. 84.
[98] See H. Syse, 2003, p. 109.
[99] A. Krishnan, 2009, p. 150; P. Singer, 2010, p.324 ; Human Rights Watch, 2012, p. 39–40; M. Ekelhof and M. Struyk, 2014, p. 10; P. Asaro, 2012, p. 692.
[100] Charter of the United Nations, Articles 2(4), 51, and 42. The 'last resort' requirement is an integral part of the international law of *jus ad bellum*. The so-called responsibility to protect is still not an established exception from the prohibition of the use of force, as it must be authorized by a Chapter VII UN Security Council Resolution. See J. Nahem and K. Egeland, 2014.
[101] The Rome Statute of the ICC did not provide a definition of the crime of aggression, but after the review conference in Kampala, Article 8 *bis* (not yet in force) defines the crime of aggression as the 'planning, preparation, initiation or execution, by a person in a position effectively to exercise control over or to direct the political or military action of a State, of an act of aggression which, by its character, gravity and scale, constitutes a manifest violation of the Charter of the United Nations' (Resolution RC/Res. 6, 11.06.2010). Thus, the crime of aggression could lead to individual criminal liability.

lives. Higher costs, both in human and financial terms, typically turn the people and electorates against their war-waging rulers. Fear of popular opposition and potential electoral defeat, then, may constrain leaders before initiating a costly war.[102]

AWSs arguably affect the threshold of the viability of the use of force through coaxing the second and, more obviously, the third of these mechanisms. Beginning in reverse order, we may expect that the more costly a government expects a war to be, the more careful it will be in avoiding it.[103] I shall in the following section first discuss the direct cost of war in terms of human lives and financial expenses, before briefly discussing the suggestion that AWSs may impact on the way in which we all, and more crucially, state leaders, think about war.

AWSs are *unmanned* systems and thus relieve the soldiers who would otherwise have manned the system or weapon of risk of harm. Current drones and other long-distance weapons display the same dynamic, whereby the operator of the weapon is located in a safe distance from the battlefield. It has been suggested by several commentators that the US 'targeted' killings on Pakistani, Yemeni, and Somali territory, killing around 3000 people by the most conservative accounts,[104] would not have been carried out in the absence of drone technology (in other words taking the use of force from 'no resort' to 'resort').[105]

While I have disputed the utility of the concept of 'full autonomy', it may make sense to call autonomous weapons 'fully unmanned'.[106] For as it happens, drone operators are not completely exempted from risk. The analogy between drone warfare and video games has often been made, but evidence

---

[102] Sometimes, the oposite may be true. H. Norpoth and A.H. Sidman (2007, p. 180), for example, argue that Britain's War in the Falklands in 1982 and the Anglo-American invasion of Iraq in 2003 secured re-election for Margaret Thatcher and George W. Bush, respectively

[103] S. Van Evera, 1999, p. 30.

[104] Bureau of Investigative Journalism, 2014.

[105] Ignatieff, 2012; J. Scahill, 2009. It is, of course, possible to make the case that the drone strikes are morally justified, and that they have made the world as a whole safer from 'terrorists'. This position has, however, been forcefully contradicted by a number of commentators, e.g. D. Kilcullen and A.M. Exum, 2009.

[106] A remote-controlled drone, then, would be 'physically unmanned'.

suggests that drone operators do not take as lightly on taking lives as some commentators have claimed.[107] According to a study conducted by the Pentagon, drone pilots were actually slightly *more* likely to get a psychiatric diagnosis than pilots of manned fighter aircraft. (The difference between the two groups was, however, not statistically significant.)[108]

AWSs would not suffer mental illness. They would thus free soldiers of both physical *and* mental risk, *ceteris paribus* decreasing the suffering associated with a war. On the other hand, the effect of a moderate decrease in mental illness among veterans on the threshold to use force *ad bellum* is probably small. Physical risk – body bags – likely have a much stronger effect. If so, the difference between AWSs and other unmanned systems such as remote-controlled drones should be negligible.

As noted, the financial cost of war is also an important determinant of the viability of war. Although the current research phase of AWSs is expensive, the complementing and replacement of human soldiers with robots is expected to reduce military expenditures in the long term. In terms of direct cost, the commonly used armed RQ-1 Predator drone cost about 4.5 million USD in 2009 – a thirtieth of the cost of an F-16 fighter jet.[109] In 2012, Human Rights Watch reported the cost of the SGR-1 sentry robots in use at the border between North and South Korea at no more than 200,000 USD.[110]

While the point about physical risk is not effected by the distinction between 'in the loop' and 'out of the loop', AWSs would by definition involve less (human) soldiers or staff in supervisory or controlling functions than remote-controlled systems, thus lowering expenditures to salaries. AWSs would, obviously, furthermore not need to be compensated for the loss of limbs or mental functions beyond mechanical repair, nor file for retirement pension or health care, whether physical or psychiatric. In the United States, veterans are eligible for annuities from the Department of Defense after 20 years of

[107] See M. Coeckelbergh, 2013, p. 93.
[108] J.L. Otto and B.J. Webber, 2013, p. 6.
[109] M.E. O'Connell, 2012, p. 267.
[110] Human Rights Watch, 2012, p. 13.

service. In 2012, 2.3 million retirees and surviving relatives received about 52 billion USD from the federal budget, a near 50 per cent increase since 2002. When health care expenditures and salaries to active personnel are included, compensation to former and present personnel eats up half the US defence budget.[111] It is, however, doubtful whether a lower wage-bill in itself lowers the threshold to use force. The interesting question would be whether actual war fighting would be cheaper with robots, not just maintaining an army.

The other way in which AWSs might arguably lower the threshold to go to war is a more indirect way: through changing the beliefs or values of electorates and political leaders regardless of whether the actual dynamic of war has changed. The attitude of leaders and the public towards war is clearly time and space contingent, and the role of technology is commonly held to have a role to play in this variability. Technology has often been cited as the solution to making wars cleaner and more humane. There is a common 'faith that technology and its energetic application can fix *anything*'.[112] According to Stephen Van Evera, state leaders commonly misunderstand or miscalculate the effect of new technologies and tactics. For example, in the 1970s, US Defense Secretary James Schlesinger suggested that a major nuclear exchange between the Soviet Union and the United States might lead to as few as 15,000 to 25,000 casualties. Today, these numbers are regarded as vastly underestimated.[113] Having introduced the machine gun, railway, and cannons with higher precision, European states plunged into the twentieth century in anticipation of a 'brisk and merry war'.[114]

In addition to their effects being miscalculated, AWSs may conceivably also change leaders' and publics' attitude towards war by creating a mental, and indeed physical, space between themselves and the battlefield. In *Just and*

---

[111] L. Montgomery, 2013. The budgetary crisis and dead-lock of the two congressional parties in the United States in 2013 was to large parts resolved through an agreement to cut pensions for military retirees.
[112] R. Johnson, 2002, p. 218.
[113] S. Van Evera, 1999, p. 32.
[114] The infamous words of a German newspaper in 1914, the first year of the First World War. See S. Van Evera, 1999, p. 33. See also W.J. Thies, 2009, pp. 63–70.

Unjust Wars, Michael Walzer suggests that '[u]ntil wars are really fought with pawns, inanimate objects and not human beings, warfare cannot be isolated from moral life'.[115] AWSs present such inanimate objects. Perhaps some indeed believe that warfare can be isolated from moral life when AWSs and not human soldiers are employed. By this line of reasoning, the use of force could be 'amoralizied' in people's minds (see Chapter 6).

A third way AWSs may lower the threshold of war is if the technology is huckstered as being, for example, more discriminate or 'ethical' than it really is. It is possible that political leaders could attempt to trick their people to get 'on board' with a war-effort. In this context, we may see AWSs as a part of a larger pattern by which political leaders attempt to legitimize and reinvigorate the use of force in the eyes of the public. While the apparent moral absoluteness of the Second World War legitimized the war's extreme cost, the Vietnam War caused moral outrage in the American public, and in other countries as well. Both American and Vietnamese casualties were perceived as far too high in relation to the danger of a communist regime in Vietnam.

According to Engels and Saas, the on-going 'War on Terror' meets acquiescence from the US public in part as a result of the rhetoric used by the presidential administration and the Pentagon, aiming to 'disempower citizens by cultivating passivity'. Acquiescent rhetorics facilitate conflict by avoiding critical inquiry and deliberation, thus 'anathema to rhetoric's nobler, democratic ends'.[116] The rhetorical woolliness could be seen in association with the use of AWSs, where the war similarly is in a sense removed or concealed from regular people (on at least *one* side of the conflict, that is). Recent wars involving North American and Western European states have displayed an impressive use and range of euphemisms by military officials and responsible politicians in order to 'obfuscate as far as possible the reality of what they do'.[117] In this discourse, 'collateral damage'[118] means the

[115] M. Walzer, 2006, p. 64.
[116] J. Engels and W.O. Saas, 2013, p. 231.
[117] R. Chambers, 2003, p. 179.

unintentional killing of civilians, 'surgical strike' practically any bombing campaign, 'enhanced interrogation' means torture, and 'neutralize' means to kill. NATO states have as far as possible avoided labelling their activities in Afghanistan, Iraq, and Libya 'war' or 'armed conflict', rather preferring the term 'operation'.[119] The new war rhetoric also includes a few misfired(!) tautologies and pleonasms, such as 'selective targeting' and 'coalition of the willing', which clumsily, and we must assume unintentionally, call into question the meaning of targeting and of forming coalitions.[120]

The hucksterism suggested in the paragraphs above should clearly be differentiated from factors that objectively lower the barriers to use force.[121] I suggest that AWSs does indeed affect the threshold to use force. Most importantly, they relieve soldiers of physical risk. On the other hand, so do bombers and remote-controlled drones. Thus, in order heighten the threshold to resort to the use force; one would have to include all riskless weapons in a prohibition. An effect unique to autonomous systems would be their financial cost – be it higher or lower than the alternatives.

Let us now revisit the slippery slope argument for AWSs and *jus ad bellum*: Is the development and use of AWSs morally wrong because it potentially lowers the threshold to go to war? Is the slippery slope argument a ground for the prohibition of AWSs?

The core of the slippery slope argument seems to be that AWSs both directly, and indirectly through miscalculation or hucksterism, could make war

---

[118] The euphemism 'collateral damage' first appeared in the Vietnam War. See R. Chambers, 2003, p. 176.

[119] For example, Norwegian authorities have refused to call their activities in Afghanistan 'war' of armed 'conflict', but have nonetheless rewarded soldiers with decorations only available to 'civilians or military personnel who *in war or armed conflict* in a particularly outstanding way have distinguished themselves' (emphasis added). My translation of 'sivile og militære som under krig eller væpnet konflikt på særlig fremragende måte har utmerket seg'. See Krigskorset med sverd, statutter, 2012, Paragraph 1.

[120] See R. Chambers, 2003, pp. 178–79. The war in Iraq also saw the emergence of the concept of 'embedding' journalists into the armed forces to provide news. Jokes about the media being 'in bed with' the military were obviously irresistible, so in this case, the euphemism failed to serve its (by critics alleged) purpose.

[121] R. Chambers, 2003, pp. 176, 179.

more palatable.[122] As war inevitably leads to harm, lower thresholds to use force can only be bad. The underlying maxim appears to be:

(1) Wars are bad.

(2) In order to avoid them, *one should not allow techniques that make war more agreeable.*

However, the positive reformulation of this maxim reveals that while the argument may be internally consistent, it is unreasonable and contradictory to its purpose of avoiding harm:

(1) Wars are bad.

(2) In order to avoid them, *one should take measures to make war less agreeable,* i.e. *more brutal and costly*.

Clearly we should neither avoid medicine nor hospitals, nor the study of strategy, nor taking precautions. Clearly we should not try to make war as brutal as possible, using, for example, inefficient weapons causing superfluous injury. There is, it seems, a slippery slope to the slippery slope argument. The logical extension of the argument reveals its inconsistency, as it appears to bind its subscriber to the implication that humanitarian law, precision weapons, armour, and chivalry are bad things.

The just war tradition solves the problem of distinguishing between fair and unfair means of making war less brutal by focusing on the intention and cause of the decision-maker, rather than on the material conditions of the threshold to resort to the use of force. As Augustine opined, it is the 'injustice of the opposing side that lays on the wise man the duty of waging wars'.[123] According to the just war tradition, the technology *per se* should have little to do with the initiation of the use of force Technologies facilitating morally

[122] See R.C. Arkin, 2008, p. 7.
[123] Augustine, 2003, p. 862 (XIX, 7). The criterion of 'last resort' could be viewed as a sub-criterion, or specification, of the requirement of a 'just cause'.

reprehensible acts should be met by making leaders more accountable and prudent, not by banning technology.[124]

The distinction between conscious hucksterism and the objective threshold to use force, however, *is* morally important. Intentionally misleading or lying to the public would clearly be morally reprehensible according to the just war tradition.[125] Yet, when it comes to the actual threshold, the *ad bellum* slippery slope argument does not appear to constitute a persuasive ground for the prohibition of, or even hostility to, AWSs.

3.1.2 The Limitations on the Use of Force *in Bello*

The slippery slope argument has not been made as vocally for the level of *jus in bello* as it has for *ad bellum*. Nevertheless, it is plausible that the argument has in fact more thrust for the conduct of hostilities. As noted above, the possibility of AWSs being better at adhering to the laws of war does not necessarily imply that less harm will be caused. Indeed, there is large room for violent acts under international humanitarian law. IHL does not prohibit the use of force, it merely limits it in certain circumstances. Hence, it is possible for AWSs to be better than humans at following the rules for conduct of hostilities without thereby causing less harm overall.

According to some commentators, the use of remote-controlled drones produces an emotional distance between the operator and his or her target, a 'video-game style interface', which may 'negatively influence the moral decision making of the human operator'.[126]

Conversely, Mark Coeckelbergh finds that surveillance technologies such as drone cameras, have made possible an 'empathic bridging' by which the enemy or target is 're-humanized, re-faced, and re-embodied'.[127] In contrast to traditional artillery and aerial bombing, drones may actually *increase* the

---

[124] The same argument would apply to the (illegal) surveillence facilitated by the new information technologies.
[125] Augustine, 2003, p. 199 (V.12).
[126] K.D. Stephan *et al*., 2012, p. 1763. See also D. Gregory, 2012, p. 197; P. Alston, 2010, p. 25.
[127] M. Coeckelbergh, 2013, p. 87.

threshold to use force *in bello*.[128] AWSs, on the other hand, will by definition function without a human in the loop, so the 'emphatic bridging' made possible by drone cameras will be lost. As noted above, drone pilots have displayed an incidence of psychiatric disorders as high as that of pilots of manned aircraft.[129]

Another concern, discussed more thoroughly in the next chapter, is that the use of AWSs may have a built-in escalatory mechanism. If AWSs were programmed to win the battles in which they were to be employed – which is, surely, a reasonable assumption – and their operationalization of the proportionality rule in IHL was a graduated response to whatever the opponent was doing, they could potentially influence the overall course of the conflict toward more harm being inflicted.

3.2 Redistribution of Harm?

Regardless of whether or not AWSs impacts the threshold to use force, the use of so-called riskless tactics and technologies leaves the victim of such tactics no choice but to remain passive or to resort to unconventional methods. More plainly: the use of unmanned systems may lead to a surge of terrorism.[130] However, such a mechanism is the case with all 'riskless' tactics, not just with AWSs, for which the novelty is the targeting technique, not the lack of risk. On the other hand, we might imagine that AWSs are perceived as more cowardly or disrespectful than other 'riskless' weapons, thus triggering a more intense and desperate opposition by the people against whom they are used.

Judging by the growing opposition to and awareness of AWSs, this appears to be a legitimate concern. There is *something* about AWSs that generates opposition.[131] In this context, then, the use of AWSs could generate negative consequences in two ways: First, it could produce more harm by

---

[128] One drone pilot explained that 'I can look at their faces. I watch them for hours, see these guys playing with their kids and wives'. It is 'not PlayStation'. See H. Gye, 2012.
[129] J.L. Otto and B.J. Webber, 2013, p. 6.
[130] See J.C. Galliott, 2012a; J.C. Galliott, 2012b, C.G. Kels, 2012; Human Rights Watch, 2012, p. 39.
[131] E.g. B.J. Strawser, 2010, p. 350.

motivating more violent opposition, and second, by redistributing harm from soldiers to civilians.

As an objection, however, such a Consequentialist position does not appear to be a strong argument for the general and total elimination of AWSs. Rather, it is a factual proposition and a caution to the people considering their use. With a view to a potential legal ban, there would furthermore probably be too many causal links involved between the deployment of AWSs and their negative consequences – potentially with improved adherence to IHL as an intermediate step.[132] The prohibitions of anti-personnel landmines through the Mine Ban Treaty and of cluster munitions through the Convention on Cluster Munitions on the other hand, were achieved only after long campaigns in which the adverse humanitarian consequences of the weapons were shown.[133] Terrorist attacks and other indiscriminate methods are, moreover, already illegal under existing international law.[134] It appears again that the problem is more with practices than with technology.

3.3 Dictators and Proliferation

Further down the slippery slope it has been suggested that AWSs may facilitate dictatorship.

> [F]rom the perspective of a dictator, fully autonomous weapons would be the perfect tool of repression, removing the possibility that human soldiers might rebel if ordered to fire on their own people.[135]

A central insight from the trials of war criminals after the Second World War was that individuals have an obligation to *disobey* orders in the event of the latter conflicting with elementary humanitarian considerations and morality. This has long been held as valid for *jus in bello* – that soldiers should refuse to commit internationally wrongful acts like war crimes and genocide – and has

[132] R.C. Arkin, 2008, p.
[133] C.H. Ruge, 2007, pp. 16–7. Mines Action Canada, 2009, pp. 2–9.
[134] E.g. API, Article 48; Rome Statute, Article 8(2)(b)(xx).
[135] B. Docherty, 2012. See also J. Hendler, 2013; M. Ekelhof and M. Struyk, 2014, p. 10.

lately been a subject also of *jus ad bellum* – that soldiers should refuse fighting in unjust wars.[136] Whether AWSs *can* be used at breach with international law remains to be seen – Ron Arkin, for example, proposes to make AWSs technically incapable of injuring civilians[137] – but it does not appear unreasonable to suspect that AWSs could be the perfect weapon for dictators without much popular support. After the Second World War, Arendt saw the increasing automation as a scary development, contributing technologies beside which the German Führer's gas chambers look like 'an evil child's fumbling toys'.[138] In his essay 'You and the Atomic Bomb', George Orwell similarly claims that while simple weapons favour revolutions and popular power, expensive and technologically advanced weapons and military equipment favour dictators and totalitarianism.[139] It seems to me that while this argument is liable to the counter-argument that it would not be the technology itself that was ethically questionable, but the practices of the dictator, it does have some merit.

Finally, some commentators have claimed that once one or a few countries have acquired the technology, the weapon systems will proliferate across the world.[140] John Canning of the US Naval Surface Warfare Center has postulated that the weapon systems in question are likely to be proliferated to a number of countries, some of them 'without the same level of safeguards that we might build in'.[141] While one may well dispute whether the United States has in fact built in many 'safeguards' for example in their drone programme at all, Mr Canning may have a point. Again, however, the argument is not really a moral one. Rather, it is a factual proposition a caution: The merit of the

---

[136] J. Wolfendale, 2009. It must be noted that such a view – whereby individual soldiers have ad bellum duties – would challange Walzer's moral equality of soldiers. See also I. Primoratz, 2002. For a criticism of the doctrine of the moral equality of soldiers see J. McMahan, 2009, 2010.
[137] More specifically, Arkin propposes to programme an 'ethical governor: which suppresses, restricts or transforms any lethal behaviour […] produced by the existing architecture so that it must fall within $P_{permissible}$ [eg. the rules governing the conduct of hostilities].' R.C. Arkin, 2008, p. 20.
[138] H. Arendt, 2006, p. 273.
[139] G. Orwell, 1968, pp 7–8. ('You and the Atomic Bomb' was also the essay in which the term 'cold war' was coined.)
[140] M. Ekelhof and M. Struyk, 2014, p. 25.
[141] J. Canning, 2005, p. 32.

argument depends on the morality of deploying AWSs in the first place. If the use of AWSs will lead to improved adherence to IHL, their proliferation would arguably be a *good* thing.

4 Autonomous Weapon Systems Under the Rules for the Conduct of Hostilities

4.1 Introductory Remarks, and the Limitation on the Choice of Means and Methods in Warfare

A second overarching argument in the Campaign against AWSs' case is their claim that AWSs will not be able to adhere to international humanitarian law. Thorough examinations have, however, been few and far between. Most often, authors content themselves to suggesting that programming AWSs with algorithms that would make them adhere to international would be difficult or impossible.[142]

4.1.1 Why IHL?

As most of the academic literature on AWSs, I shall in the following primarily discuss IHL. One of the most interesting developments in the Campaign to Stop Killer Robot's advocacy was the informal expert discussions in the CCW – an IHL instrument – in May. One should, however, note that IHL would not necessarily be the only applicable legal regime. When used as part of a police force in law enforcement, for example, the use of AWSs would be subject to domestic law and human rights law (HRL).

While IHL clearly applies to war, the role of human rights law (HRL) in armed conflict has been a matter of some contention.[143] There is now a growing consensus that human rights law (HRL) is not displaced by, but complementary to, IHL as a regulatory framework of armed conflict, and that the application of one or the other to a large degree depends upon the context in which the

[142] E.g. M. Ekelhof and M. Struyk, 2014, pp. 11–5; Human Righs Watch, 2012, pp. 30–4; W. Wallach and C. Allen, 2013, p. 125; P. Asaro, 2012, pp. 695–700; H.Y. Liu, 2013, p. 640–3; M.N. Schmitt, 2013; N. Sharkey, 2007, pp. 124–5, 2008, p. 16–7; R. Tonkens, 2013, pp. 153–6; A.M. Johnson and S. Axinn, 2013, pp. 130–3; R.M. O'Meara, 2012, pp. 160–3.
[143] See e.g. G. Corn, 2010, p. 53; C. Tomuschat, 2010, pp. 20–1.

violence is taking place.[144] Moreover, as IHL as a whole also takes a somewhat more sinister view of issues of compliance and practicality, in that respect it is more in line with my design choice of employing non-ideal theory (see Chapter 2). Hence, the following discussion primarily leans on IHL instruments and their interpretation.[145]

My opting to investigate the legality of AWSs in international armed conflicts, then, excludes questions of their use in so-called armed violence or criminal conduct regulated by law enforcement rules and HRL. Furthermore, my focus on *international* armed conflicts to a certain extent further reduces the universe for which the inferences drawn in this chapter are supposed to hold, as there are separate rules for international and non-international armed conflicts. Whereas the label 'international armed conflict' (IACs) denotes conflicts between sovereign states, 'non-international armed conflicts' (NIACs) covers conflicts between states or governments and armed opposition groups within that state. Borderline situations may arise when ideal type non-international armed conflicts are complicated by third-party intervention or cross-border operations.[146]

In a historical perspective, the laws of international armed conflict have a longer tradition than the laws of non-international, or internal, armed conflicts, and their provisions are generally more extensive.[147] Following the

---

[144] For IHL to apply, the act must in some sense be a part of, or contribution to, the central conflict.
[145] IHL is the standard framework for the regulation of means and methods of armed conflict. To a greater degree than HRL, IHL recognizes military necessity as a counterweight to humanitarian principles. This makes IHL a more permissive legal regime than HRL and law enforcement rules regarding the use of force. In practice, however, human rights, as part of the rules for law enforcement, are not unequivocally less permissive than IHL regarding the legal scope of arms use. For example, typical gaseous riot control agents, such as tear gas, are prohibited by the Chemical Weapons Convention under IHL, but not by law enforcement rules. The same goes for expanding bullets (Hague Convention III, 1899). This is due to the fact that while IHL is a symmetric system viewing belligerent parties as at the outset moral equals, HRL and law enforcement are asymmetric, aiming to direct harm towards the 'guilty'. This allows for harsher means of protecting bystanders and removing threats. Under law enforcement rules, the distinction between civilians and combatants present in IHL does not exist, as all involved are *citizens* under the law, not *combatants* or *civilians*. Hence, assertions that IHL 'is, in effect, less protective for civilians than a law enforcement framework which includes IHRL' (H. Matthews, 2013, p. 634), are essentially nonsensical statements regarding the false distinction between civilians and combatants in law enforcement, displaying a damaging lack of nuance.
[146] D. Feck, 2008, pp. 605–6.
[147] D. Fleck, 2008, p. 612. The more permissive outlook of the rules regulating non-international armed conflict should not come as a surprise, as international law is created by states, who themselves have

US doctrine of Air–Sea Battle, I believe the prospect of AWS being used is a real one, and perhaps most likely in an IAC featuring high-tech states, as states would be much more likely than non-state actors to possess the type of equipment necessary for the use of autonomous weaponry (e.g. sophisticated cyber capabilities).[148] However, many of the core principles of IHL are common to both international and non-international armed conflicts, (although their interpretation arguably vary across the regimes). In that respect, much of the discussion below will have validity for both IACs and NIACs.

4.1.2 Background

Whether an armed conflict is of an international or non-international nature, the liberty of the parties to the conflict to choose means and methods of warfare is not unlimited.[149] This is a core principle of IHL, often called the law of armed conflict (LOAC), or *jus in bello*.[150] This foundation is considered to be customary law, and applicable to both international armed conflicts (IACs) and non-international armed conflicts (NIACs).[151]

As is the norm in international law, treaties and conventions are generally only binding for the states that have ratified them. However, many of the most important rules and principles of IHL are considered customary international law,[152] defined by the Statute for the International Court of Justice (ICJ) as 'evidence of a general practice accepted as law'.[153] All the World's States, moreover, have ratified the four Geneva Conventions. In its *Nuclear Weapons Advisory Opinion* (1996), the ICJ asserted of Additional Protocol I to the Geneva Conventions of 1949 – arguably the most important instrument

---

few interests in extending the rights of potential enemies within. The protection of states from interference in their internal affairs made its breakthrough in Münster in 1648, and is arguably one of the strongest in all of international law (see G. Nystuen and K. Egeland, forthcoming, 2014).

[148] Air–Sea Battle Office, 2013, p. 2; N. Sharkey, 2013; W. Pincus, 2012.

[149] Annex to Convention (IV) Respecting the Laws and Customs of War on Land: Regulations Concerning the Laws and Customs of War on Land, the Hague, Article 22; API, Article 35(1).

[150] S. Oeter, 2008, p. 126.

[151] D. Fleck, 2008, p. 613–4.

[152] C. Greenwood, 2008, p. 33; D. Fleck, 2008, p. 608.

[153] Statute for the International Court of Justice, 1945, Article 38(1)(b). Louise Doswald-Beck defines customary law more soberly, as 'the normal conduct and behavior expected of States' (1997).

codifying the rules for the conduct of hostilities – that 'the Court recalls that all States are bound by those rules in Additional Protocol 1 which, when adopted, were merely the expression of the pre-existing customary law, such as the Martens Clause'.[154] The Court did not further specify which rules they had in mind, but labelled the rules of distinction and avoiding unnecessary suffering 'cardinal'. One may accordingly safely assume that they are to be covered by the Court's assertion.[155] According to the International Committee of the Red Cross' Study on customary international law from 2005, all of the core conduct of hostilities rules – distinction between civilians and combatants (including the prohibition against indiscriminate weapons), proportionality between ends and means, taking precautions in attack, and avoiding unnecessary suffering – are customary.[156] In practice, this implies that the United States, for example, is obliged to follow the core principles of distinction, proportionality, and of taking precautions in attacks,[157] even though it has not ratified Additional Protocol I, in which these principles are codified.[158]

The core rules governing the conduct of hostilities are codified in API, the most important of which are (1) distinction, (2) proportionality, (3) precautions, and (4) avoidance of unnecessary suffering. [159] Numbers (2) and (3) are sometimes considered as operationalizations of the 'cardinal' principle of distinction. In addition to these, the so-called Martens Clause provides that the 'principles of humanity' and 'dictates of public conscience' shall be

---

[154] *Legality of the Threat or Use of Nuclear Weapons*, International Court of Justice Advisory Opinion, 08.07.1996, Paragraph 84.
[155] *Legality of the Threat or Use of Nuclear Weapons*, International Court of Justice Advisory Opinion, 08.07.1996, Paragraph 78.
[156] J.M. Henckaerts, 2005, pp. 87, 93.
[157] API, Articles 48, 50, 51, and 57.
[158] C. Greenwood, 2008, p. 33; API, Article 37. The four Geneva Conventions of 1949 are ratified by 195 parties, including all UN member states. Thus, the question of whether these conventions – or which of their principles – may be regarded as customary law or not is in effect irrelevant.
[159] M.N. Schmitt, 2010, pp. 803–4. According to the legal scholar and International Court of Justice judge Cristopher Greenwood, IHL may be distilled into three principles. Acts of war are only permissible if, and only if, they are (1) directed against military objectives, (2) are not likely to cause unnecessary suffering, and (3) if they are not perfidious (C. Greenwood, 2008, p. 35). Greenwood explains the prohibition of perfidy as meaning that the parties are obligated not to use 'treacherous methods and means of warfare' (API, Article 37; C. Greenwood, 2008, p. 37). Generally, it is taken to mean that it is illegal to fly false colours, to mislead the enemy into believing that one is seeking truce or ceasefire only to make a surprise attack, and similar tactics, should they entail loss of life (S. Oeter, 2008, p. 227–9).

binding even in cases not covered by the Protocol. In the following section, I assess their applicability to the question of employing AWSs.

A different, but related, question that has been raised by people and organizations associated with the international campaign against 'killer robots',[160] is whether the alleged advent of AWSs call for a legal review of these weapon system's legality under IHL, as codified by Article 36 of Additional Protocol I:

> In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.[161]

Few such reviews appear to have been made, no doubt because these systems are still on the drawing board. This thesis may in part be seen as an attempt at addressing this paucity. As AWSs are arguably not yet in existence, it is obviously difficult to evaluate their adherence to international law. I shall in the following rely on current technology, the direction of the development, and the suggested solutions for the future.

4.2 Distinction and Inherently Indiscriminate Weapons

Parties to an armed conflict may only direct attacks against military objectives. According to Stefan Oeter, the norm of distinction, or discrimination,[162] between combatants and civilians is not only to be considered customary international law, but arguably 'constitutes one of the few peremptory norms of humanitarian law, and accordingly also part of "*ius cogens*"'.[163] Article 48 of Additional Protocol I provides that the state parties must always be able to

[160] E.g. ICRC, 2013; Reaching Critical Will, 2013. One of the NGOs campaigning against AWSs is amply named 'Article 36'.
[161] API, Article 36.
[162] The terms are often considered synonymous. See S. Casey-Maslen, forthcoming, 2014.
[163] S. Oeter, 2008, p. 136. See also S. Casey-Malsen, forthcoming 2014. A premptory, or *jus cogens*, norm, is an imperative norm of higher rank than ordinary rules (see A. Cassese, 2001, pp. 138-39).

distinguish between both military personnel and civilians, and between military and civilian objects and infrastructure:

> In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.[164]

The status of persons as civilian, combatant, prisoner of war, interned, or soldiers *hors de combat* determines how people affected by armed conflict should be treated.[165] Attaining the technical capabilities demanded to be able to make these distinctions would be a necessary (but by itself insufficient) condition for the lawful deployment of AWSs. The first, and fundamental question, then, is how, and if, AWSs would be capable of discriminating between (lawful) targets and persons enjoying non-combatant immunity. Before investigating further, we may note the three ways in which a weapon may be used unlawfully:

First, the use of some weapons is explicitly prohibited by specific legal instruments. For example, the Chemical Weapons Convention (CWC) prohibits basically any dealings with chemical weapons, including use and stockpiling.[166] Specifically prohibited weapons, such as the noted chemical weapons, bacteriological weapons, blinding lasers, land mines, and cluster munitions, have been prohibited on the grounds that they are either of a nature to cause unnecessary suffering or cannot be used in accordance with the norm of distinction (or both). There is as of yet no specific treaty law for AWSs.[167]

Second, even though a weapon is not specifically prohibited by a treaty, it may still be categorically unlawful to use if it *inherently* conflicts with the

---

[164] API, Article 48.
[165] K. Ipsen, 2008, p. 78.
[166] Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction, Articles 1(1).
[167] The process toward a potential ban or regulation of AWSs was begun in May 2014, with discussions by an expert diplomatic group in the CCW.

principles of distinction and avoidance of unnecessary suffering.[168] The use of such inherently indiscriminate weapons is prohibited regardless of any specific instrument.[169]

Third, a weapon may be *used* unlawfully without being inherently indiscriminate or in any other way at breach with international law. In its advisory opinion on Nuclear Weapons from 1996, the ICJ maintained that nuclear weapons were not illegal *per se*, but would be difficult to use without violating the principles of distinction and avoidance of unnecessary suffering.[170] At the most basic level, however, even standard rifles have a maximum range of precision.

In the case of AWSs, the second of these grounds of unlawfulness appears the most interesting: Are AWSs inherently indiscriminate? The *UK Manual on the Law of Armed Conflict* identifies the issue of inherently indiscriminate weapons as follows:

> It is prohibited to employ weapons which *cannot be directed* at a specific military objective or the *effects of which cannot be limited* as required by Additional Protocol I and consequently are of a nature to strike military objectives and civilians or civilian objects without distinction.[171]

From this we may identify two ways a weapon may be inherently indiscriminate. On the one hand, some weapons cannot be targeted against a specific target, and are thus prone to be indiscriminate on impact. On the other hand, a weapon may have uncontrollable effects, such as incendiary weapons designed to start fires. The following two sections are devoted to each of these grounds of unlawfulness.

---

[168] API, Articles 35 and 48. See also L. Doswald-Beck, 1997.
[169] API, Article 51. See also L. Doswald-Beck, 1997.
[170] ICJ, *Legality of the Threat and Use of Nuclear Weapons*, 1996, Paragraph 95.
[171] The Joint Service Manual of the Law of Armed Conflict (UK Manual), 2004, Paragraph 6.4. Emphasis added.

4.2.1 Indiscriminate Targeting?

As noted, '[i]t is prohibited to employ weapons which cannot be directed at a specific military objective'.[172] In the Nuclear Weapons Advisory opinion, the ICJ postulated that '[s]tates must […] never use weapons that are incapable of distinguishing between civilian and military targets.'[173] To this, Louise Doswald-Beck notes that weapons, obviously, cannot *themselves* distinguish between civilians and combatants:

> It remains to be seen what precisely the Court meant by 'incapable of distinguishing between civilian and military targets'. It is obvious that a weapon, being an inanimate object, cannot itself make such a distinction, for this process requires thought.[174]

In the context of AWSs, we are immediately faced with a challenge: In what sense is an AWS directed, and in what sense is it a weapon?

As I have defined it above, an AWS is 'a weapon that can independently select and attack targets.'[175] It would be impossible for humans to directly engage 'a specific military objective' by means of an AWS. This would contradict the definition of AWSs as acting 'independently'. Instead, the targeting is performed by the system itself. The AWS identifies a target independently, and then releases the weapon. The actual weapon could take almost any conceivable form, being anything from a machine gun firing bullets to an intercontinental ballistic missile. The novelty introduced by AWSs is not the weapon, but the targeting technique.

In contrast to the *UK Manual* and the Nuclear Weapons Advisory Opinion, which both describe unlawful indiscriminate *weapons*, API does not

---

[172] The Joint Service Manual of the Law of Armed Conflict (UK Manual), 2004, Paragraph 6.4. See also API, Article 51.
[173] *Legality of the Threat or Use of Nuclear Weapons*, International Court of Justice Advisory Opinion, 08.07.1996, Paragraph 78. The Court's use of the phrase 'civilian target' is also troubling. According to the principle of non-combatant immunity, civilians cannot be 'targets'. See S. Casey-Maslen, 2014 (forthcoming).
[174] L. Doswald-Beck, 1997.
[175] ICRC, 2014b, p. 1.

prohibit indiscriminate weapons, but indiscriminate *attacks*.[176] In Article 51 (4) (b) such attacks are defined as '[t]hose which employ a method or means of combat which cannot be directed against a specific military objective'.[177] This broadens the class of phenomena covered considerably.

Furthermore, the definition of AWSs – as weapons that can 'independently select and attack targets'[178] – implies that the weapons *are* discriminate. A 'target' can only be understood as a military object or enemy soldier. The whole point of the rule of distinction is precisely that a civilian is *not* a target. Thus, given many civil society actors' and states' claim that AWSs will not be able to discriminate between soldiers and civilians, their inclusion of the word 'target' in the definition of AWSs is puzzling.[179] For example, the Dutch NGO Pax for Peace, founding member of the Stop Killer Robots campaign, defines AWSs as 'weapons that can select and engage targets without human intervention.'[180] On the other hand, they claim that AWSs 'cannot be programmed in such a way that they will be able to make sound decisions about who is a combatant and who is a civilian.'[181] Such statements come dangerously close to tautology. Several of the general statements at the CCW expert discussions in May 2014 displayed the same confusion.[182] Hence, if the word 'target' is to be maintained in the definition, the question is less about whether AWSs can distinguish between combatants and non-combatants, and more about whether it is possible to create AWSs at all.

---

[176] API, Article 51(4).
[177] API, Article 51(4)(b).
[178] ICRC, 2014b, p. 1.
[179] E.g. Campaign to Stop Killer Robots, 2014; Human Rights Watch, 2012, p. 2; US Department of Defense Directive, 2012, p. 14). It is
[180] M. Ekelhof and M. Struyk, 2014, p. 4.
[181] M. Ekelhof and M. Struyk, 2014, p. 13.
[182] E.g. Norwegian Ministry of Foreign Affairs, 2014; Japanese Ministry of Foreign Affairs, 2014; Swiss Ministry of Foreign Affairs, 2014. It is reasonable to suspect that some NGOs and states are unaware of the legal ramifications and subtle meaning of the word 'target'.

A Note on Drone Strikes

There are important differences between AWSs and armed drones (UAVs). While many drone platforms have automated capabilities, they are remote-controlled at the crucial moment of kinetic action. An AWS, on the other hand, is capable of inflicting lethal harm without any direct human interference. Nevertheless, there are some important similarities in how they would identify targets. In the following, I discuss how so-called profile or signature strikes display important similarities to how AWSs are intended to function.[183]

A general trait of many modern conflicts is that the tactics employed by one or more of the belligerent parties make discrimination between combatants and civilians difficult to conduct in practice. Arguably, most wars since the Second World War display a wide use of so-called fourth generation warfare tactics (4GW), such as guerrilla warfare, terrorism, and immersion in the local population.[184] Parties using such tactics seldom wear uniforms, and generally do their best *not* to distinguish themselves from the civilian population (which is a requirement of IHL),[185] precisely to leech on the protection afforded to civilians. This makes it very difficult for their (for now human) adversaries to distinguish them from civilians.

An increasingly used counter-tactic, employed on a large scale in Afghanistan, Iraq, Libya, and Gaza,[186] is the use of armed drone aircraft (unmanned aerial vehicles, or UAVs). With this tactic, targets are selected through the camera of a drone, often based not on whether they are engaging in hostile activities or are identified as legitimate targets, but based on predetermined criteria, such as location, gender, or 'suspicious' looks or

[183] The issue most commonly debated when discussing drones – the question of breaching the sovereignty of other countries, for example the US in Pakistan, will not be rehearsed here, as it does not appear directly linked to AWSs and the conduct of hostilities. Rather, it is a question of *jus ad bellum*, or the initiation of war.
[184] T.X. Hammes, 2004, pp. 44–6. 4GW was, according to Hammes, invented by Mao Zedong, and can be summarized by its three stages: (1) Propaganda, (2) small-scale insurgency, alliance creation, and winning 'hearts and minds', and (3) large-scale war/insurgency.
[185] API, Article 44(3). See also H.P. Gasser, 2008, p. 239.
[186] B. Emmerson, 2013, Paragraphs 29–39.

activity.[187] In their report on drone strikes, Stanford Law School concluded that under the drone programme, 'enemies' were targeted based on 'signatures' associated with terrorist activities:

> Under Obama, the [drone] program expanded to include far more 'profile' or so-called 'signature' strikes based on a 'pattern of life' analysis. According to US authorities, these strikes target 'groups of men who bear certain signatures, or defining characteristics associated with terrorist activity, but whose identities aren't known.' […] The *Times* also reported that some in the Obama administration joke that when the CIA sees 'three guys doing jumping jacks,' they think it is a terrorist training camp.[188]

The RQ-9 Reaper, produced by General Atomics, can 'self-navigate and search out targets, but a remote operator […] makes the final decision to release the missiles'.[189] In the future it would be possible to replace the human drone operator with computer algorithms completely.[190] While there are still humans controlling the actual drone strikes, the selection of targets is not in any meaningful sense made by the drone operator, but *a priori* by predetermined criteria, or 'signatures'. Often, what is targeted is not even an actual person, but the SIM card associated with a suspected enemy.[191] Presumably, the targeting method of AWSs would be similar, *i.e.* based on sensorial recognition of predetermined, objective criteria.[192] 'Gait recognition' is one such suggested discrimination-technique, by which persons can be identified by their looks and 'specific walk'.[193]

The qualitative difference between today's drones, with pilots or operators 'in the loop', and tomorrow's AWSs with humans 'out of the loop', has often been stressed. However, there is clearly a massive grey area between

---

[187] D. Greenfield, 2013.
[188] Stanford and NYU, 2012, pp. 12–3.
[189] N. Sharkey, 2008, pp. 14–5.
[190] A. Quince, 2013. An algorithm is a 'process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer' (OED, 2014, 'Algorithm').
[191] J. Scahill and G. Greenwald, 2014.
[192] See R.C. Arkin, 2008, p. 11.
[193] T. Spenser, 2012.

AWSs guided by artificial intelligence and remote-controlled automated weapons and platforms.

Drone strikes have led to a large number of civilian deaths; allegedly, children, people trying to rescue or tend to the wounds of drone strike victims, funeral processions, and wedding guests are among those who have been targeted.[194] Proponents of the drone strike programme claim that drones in fact provide better opportunities to distinguish between civilians and combatants than alternative means and methods. For example, the MQ-1 Predator and the MQ-9 Reaper – some of the most used armed drones – can stay in the air for up to fourteen hours, compared to a maximum of four hours of F-16 fighter jets and A-10 ground attack aircraft.[195] This allows strikes to be made on the basis of better intelligence. Compared to cruise missiles or intercontinental ballistic missiles, which are released far away from the intended target, the drone platform permits the missile to be launched much closer to the target. This gives their operators the opportunity to change their minds at a much later stage should they observe unlawful targets in the targeted area. By contrast, the cruise missiles aimed for Osama bin Laden by the United States in 1998 were programmed to hit the area in which he was presumed to be located four to six hours after firing the missiles.[196] Such attacks are clearly both bad tactics and questionable under international law.[197]

However, while the technology of the drones does allow for better distinction than many other systems, they have not always been used to that effect. Military strategists David Kilcullen and Andrew M. Exum assert that drones are essentially a tactic, or actually a piece of technology, that became a strategy. The problem is not primarily with the technology, but with how it is put to use.[198] Furthermore, while the intelligence gathering made possible by

---

[194] M. Zenko, 2013, p. 14; J. Becker and S. Shane, 2012; C. Whitlock, 2011; G. Greenwald, 2012.
[195] W. Wheeler, 2012.
[196] M. Zenko, 2013, p. 6.
[197] Note, however, that the assassination attempt at bin-Laden would not have been regulated by IHL, as there was no armed conflict between the parties involved. Rather, one would have to draw upon HRL and *jus ad bellum*.
[198] D. Kilcullen and A.M. Exum, 2009.

drones is better than that of fighter jets, it does not compare to having eyes and ears on the ground.[199] A recent study found that 'unmanned vehicles were 10 times more likely than conventional aircraft to cause civilian casualties'.[200]

Few disagree that the US drone strike policy is in need of reform, and according to R.C. Arkin, development of AWSs is precisely the way forward. He argues that AWSs could act more ethically than human beings on the battlefield, not least because they would not have human emotions, which are often responsible for rash decisions and thus unlawful acts.

> It is not my belief that unmanned systems will be able to be perfectly ethical in the battlefield, but I am convinced that they can perform more ethically than human soldiers are capable of. Unfortunately the trends in human behavior in the battlefield regarding adhering to legal and ethical requirements are questionable at best.[201]

The technology available today is by all accounts far away from a scenario in which robots would be capable of engaging legitimate targets and not civilians, with or without uniformed enemies.[202] Compare, for example, the quote directly above, to a description of Samsung's SGR-A1 sentry robot used at the 'demilitarized zone' (DMZ) between North and South Korea, one of the most sophisticated weapon systems available today. Although it can distinguish people from rocks, trees, and potentially animals, it cannot distinguish between combatants and non-combatants:

> [SGR-A1] is equipped with sophisticated color vision sensors that can identify a person entering the DMZ, even at night under only starlight illumination. Since any person entering the DMZ is automatically presumed to be an enemy, it is not necessary to separate friend from foe.[203]

---

[199] D. Kilcullen and A.M. Exum, 2009.
[200] J. Scahill and G. Greenwald, 2014.
[201] R.C. Arkin, 2008, p. 7.
[202] N. Sharkey, 2008, p. 16; P. Asaro, 2012, pp. 697–8; H.Y. Liu, 2013, p. 651; R.C. Arkin, 2008, p. 11; V. Morkevicius, 2014, pp. 3–4.
[203] P. Lin *et al.*, 2008, p. 19.

While engineers have been able to create robots that mimic human movements, and computer processors are capable of mathematical calculations by far superior to anything a human could do, creating robots with effective language functions – not just automatic recordings and speech recognizing algorithms – has been laborious and unfruitful. Acquiring language would require the capacity to interact with other agents under continuously changing circumstances, in that sense similar to the chaos of battlefield operations.[204] However, complications and difficulties in the present do not imply that robots will *never* reach the level of understanding necessary to adhere to the norm of distinction. To the contrary, the technological development is moving fast:

> Innovations, adaptations, and uses in […] robotics, are being discovered at an unprecedented rate in a culture of technological uncertainty, which provides very little time and minimal governance in order to ask the question of not can we do this, but should we do this.[205]

The question of whether robots could discriminate between combatants and non-combatants is essentially a factual one, boiling down to whether robots can be made to think and function like human beings. This is, incidentally, also how the US Department of Defense phrases it in their most recent roadmap for unmanned systems: 'In simplistic terms, the algorithms must act as the human brain does'.[206]

4.2.2 Uncontrollable Effects?

Above I have argued that future robots may be capable of discriminating between lawful and unlawful targets when using force. In this regard it is important to stress that the *weapon* will not be directed by human commanders or soldiers, but by the autonomous system or robot. If humans could in fact

---

[204] See M. Brooks, 2013. It is tempting to call attention to W.H. Auden's poem 'August 1968': 'The Ogre does what ogres can,// Deeds quite impossible for Man,// But one prize is beyond his reach,// The Ogre cannot master Speech:// About a subjugated plain,// Among its desperate and slain,// The Ogre stalks with hands on hips,// While drivel gushes from his lips' (2007, p. 304).
[205] R.M. O'Meara, 2012, p. 159.
[206] US Department of Defense, Unmanned Systems Integrated Roadmap, 2013, p. 67.

direct the weapon, the robot or system would not be autonomous, but automated or remote-controlled. The principle of distinction, however, goes beyond the immediate targeting and impact of an attack. It also concerns longer-term consequences of an attack. Article 51 of Additional Protocol I further defines indiscriminate attacks as:

> Those which employ a method or means of combat *the effects of which cannot be limited* as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction.[207]

This temporal dimension of the rule of distinction would, for example, apply to nuclear or other radiation or destruction of vital infrastructure. Again, the term 'method or means' is important. The question here is not whether the actual *weapon* is indiscriminate, but whether the autonomous system is. Hence, the question must be whether the autonomous *system* can be limited. This has both a spatial component (discussed above) and a temporal component. The temporal component – how the system would be deactivated after completing a mission – may seem trivial, but in fact, it is not, as it spills over into what in robotics is known as the 'human interference issue'.[208] At a minimum, creating simple 'on/off' buttons would be a necessity. Due to their financial cost, however, we might reasonably expect that military commanders and other decision-makers would be reluctant to leave valuable technology in the hands of the enemy. As Lin *et al*. puts it, future robots will be better than humans at assessing risk:

> We cannot trust humans to determine risks for autonomous robots, not least because we are often psychologically, emotionally, and cognitively ill-equipped to accurately understand and estimate the risk. As robots grow in their lethality, speed, and autonomy, this problem will only become more acute.[209]

---

[207] API, Article 51(4)(c). Emphasis added.
[208] M.N. Schmitt, 2013, p. 6.
[209] P. Lin *et al*., 2008, p. 71.

Nevertheless, assuming that 'on/off' buttons are created, it appears unreasonable to suggest that AWSs will be intrinsically indiscriminate or impossible to limit in their effects along the time dimension. According to Schmitt, the 'U.S. Department of Defense is exceptionally sensitive to the human interface issue'.[210] In the directive issued by the US Department of Defense in 2012, decision-makers were obliged to design 'human-machine interfaces for autonomous and semi-autonomous weapon systems to be readily understandable to trained operators, provide traceable feedback on system status, and provide clear procedures for trained operators to activate and deactivate system functions'.[211]

4.3 Proportionality and Military Necessity

The norm of distinction is not peremptory in the sense that all attacks causing civilian deaths are unlawful. The principle of distinction must constantly be weighed against military necessity, and is mediated through the norm of proportionality. In line with realist reasoning, it was once held by some writers that military necessity – winning the war or a battle – prevailed over humanitarian considerations such as avoiding civilian casualties. Today, this claim is decisively refuted. In the so-called subsequent Nuremberg Trials, for example, the United States Military Tribunal in *United States v List* ruled that 'military necessity or expediency do not justify a violation of positive rules'.[212] All civilian casualties must be measured against the importance of the operation. A small tactical advantage is not sufficient to justify large numbers of civilian casualties. Thus, the prohibition against the use of indiscriminate weapons also includes a proportionality clause:

> Indiscriminate attacks are prohibited. Indiscriminate attacks are: […] An attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof,

[210] M.N. Schmitt, 2013, p. 6.
[211] US Department of Defense, Department of Defense Directive, 2012, Paragraph 8(a)(4).
[212] Quoted in C. Greenwood, 2008, p. 38.

> which would be *excessive in relation to the concrete and direct military advantage anticipated*.[213]

Here, an even more careful judgement is called for than the more instrumental obligation to distinguish between civilians and combatants, and between active combatants and soldiers *hors de combat*, including injured and surrendered soldiers. The proportionality rule is very complex, obliging parties to assess the commensurability between means and methods on the one hand, and the military advantage anticipated on the other. In order to adhere to this rule, then, an AWS would have to have a clear understanding of when anticipated civilian harm would be excessive in relation to the military advantage gained. This would necessitate an understanding of military grand strategy, operational issues, and tactics. The AWS would furthermore have to be able to comprehend continual changes in goals and objectives, internal changes to their relative importance, and the anticipated military utility of achieving them. Ultimately, these understandings would have to be operationalized as the number of civilian casualties (and damage to the natural environment)[214] these goals are worth, assuming that the other rules for the conduct of hostilities are already met. These are very complicated tasks, that only humans are capable of today. According to Peter Asaro, '[h]uman understanding, rationality, and judgement exceed any conceivable system of fixed rules or any computational system'.[215]

The answer to the question 'what is excessive?' is, as Yoram Dinstein understatedly puts it, 'not exact science'.[216] The application of IHL requires an operationalization of abstract principles, to which there is a certain reciprocal element.

IHL is intended as a de-escalating mechanism, and it may well be disputed whether robots are able to grasp this. Proportionality can be operationalized in two ways: First, it can be determined *before* a mission how

---

[213] API, Articles 51(4) and (5)(b). Emphasis added.
[214] See L. Doswald-Beck, 1997.
[215] P. Asaro, 2012, p. 700. See also V. Morkevicius, 2014.
[216] Y. Dinstein, 2004, p. 122.

important its success is. In the case of AWSs it appears very difficult and labour intensive to re-programme the software for every mission, so while this is the standard way to operationalize the proportionality rule with human soldiers it may less viable for robots. Second, the proportionality rule can be applied as a *graduated response* to what the enemy is doing, in other words operationalized as 'the principle of minimum use of force'.[217] This appears more feasible when deploying AWSs. A major problem, however, is that this graduated response has the implication – particularly when both sides are equipped with AWSs – of being an escalatory mechanism; obviously, both sides are attempting to win the battle, which may be equally important for them. 'Minimum' does by no means imply 'minimal'. Thus, AWSs may be militarily less useful than hoped, due to their capacity to influence the conflict on a general, strategic level. Granting machines control of the use of force could potentially lead to the balance between military necessity and humanitarian considerations being shifted in favour of the former.

However, as with the discussion of the rule of distinction above, it is very difficult to assess the legality of systems that are not yet in use, and it is not clear that humans are unequivocally better than computers at calculating proportionality, or at least the number of civilian casualties resulting from any given strike. The US military in fact already uses computer software to aid their proportionality assessments.[218] Furthermore, no weapon, and arguably no means nor method of war, have been prohibited on the grounds of being inherently disproportionate.[219] Disproportion would have to be determined in

---

[217] ICRC, 2014a.
[218] The software is called 'Fast Collateral Damage' and was formerly (rather unfortunately) known as 'BugSplat'. See US Joint Forces Command, 2007.
[219] See S. Casey-Maslen, forthcoming 2014. As noted, several weapons have been banned on the grounds of being inherently indiscriminate. This also goes for the broader category of indiscriminate acts or 'methods and means', as for example acts of terror and perfidy are also prohibited under international law. Whether a weapons *platform* has ever been prohibited is a point of some controversy. Arguably, an anti-personnel landmine is not just a weapon, but also a platform. Around the turn of the twentieth century, a campaign to ban aerial bombardment from planes and balloons was launched. The campaign had some success, as the Hague Declaration (IV,1) prohibits 'for a term of five years, the launching of projectiles and explosives from balloons, or by other new methods of similar nature (Declaration (IV,1), to Prohibit, for the Term of Five Years, the Launching of Projectiles and

light of the military utility of the attack, and the actions of the AWS on the battlefield. Furthermore, one could imagine a situation in which the risk of harming civilians was limited. The proportionality rule does not protect enemy combatants. According to Dinstein, there is nothing legally wrong with a one-sided military victory:

> There is no legal fault in a one-sided military success of a belligerent party benefiting from supremacy of means and methods of warfare from pursuing a 'zero casualties policy', leaving its own combatants intact, while inflicting horrific losses on the enemy's armed forces.[220]

4.4 Precautions and Avoiding Unnecessary Suffering

Even if an AWS could properly discriminate between combatants and civilians, the effects of it could be limited in time, and it could judge the relation between military advantage and harm to civilians, its actions could still be unlawful on the grounds of breaching the rule of taking precautions in attack. Article 57 of API reads that '[i]n the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects'.[221] Commenting on this rule, Stefan Oeter explains that '[a]ttacks against military objectives shall be conducted with maximum precautions to protect the civilian population', and furthermore that '[t]he rule prohibits not only attacks against the general civilian population but also attacks likely to cause incidental loss of civilian life, injury to civilians, or damage to civilian objects which are excessive in relation to the expected advantage'.[222] Moreover, the provision specifies the following:

> An attack shall be cancelled or suspended if it becomes apparent that the objective is not a military one or is subject to special protection or that

---

Explosives from Balloons, and Other Methods of Similar Nature. The Hague, 29 July 1899; see also Constitutional Rights Foundation, 2014).
[220] Y. Dinstein, 2013, p. 80.
[221] API, Article 57(1).
[222] S. Oeter, 2008, p. 189.

> the attack may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof.[223]

Thus, belligerents are required to assess the proportionality between ends and means also during the course of the fighting. Arkin proposes that AWSs could be programmed with algorithms for 'case-based reasoning', which could presumably adjust the system's interpretation of the proportionality rule prior to a specific mission.[224] On the other hand, even this could prove too crude an application, as the importance of the mission and thus advantage anticipated will often change during the course of the fighting. According to Janina Dill, the balancing between military advantage and civilian losses implied by the proportionality requirement is especially:

> The requirement to balance without a specified mechanism for how to do so is difficult on its own. It is further complicated by disagreement over who counts as a civilian with immunity from attack and over what exactly constitutes a military advantage.[225]

Moreover, IHL also provides that unless circumstances do not permit, attacks which may affect civilians should be preceded by a warning in order for civilians to escape the area.[226] Admittedly, this would defeat the purpose of targeted killings and assassinations as in the case with the drone strikes discussed above. However, there is clearly a difference between civilian casualties when effective warning have been given and when it has not. A strategy *based on* assassinations without due warning is certainly questionable given existing IHL. On the other hand, one may reasonably hold that if technology advances far enough so that a machine can adhere to the norms of distinction and proportionality, it could also be made to take precautions.

Another concern is how one might practically go about surrendering to a robot or weapon system. This relates to the technological challenges of performing the necessary distinction between combatants and non-combatants

---

[223] API, Article 57(2)(b).
[224] R.C. Arkin, 2008, p. 64.
[225] J. Dill, 2010, p. 3.
[226] API, Article 57(2)(c).

discussed above, but goes beyond the kinetic action. The robot would, for example, have to be capable of distinguishing between active combatants and surrendered or surrendering soldiers, and others *hors de combat*.[227]

In addition to the rules of distinction, precautions, and proportionality, there is a strong rule in IHL calling combatants to use means and methods aiming at avoiding unnecessary suffering for combatants.[228] The rule of avoiding unnecessary suffering has been expressed in several legal instruments banning specific weapons. The first convention banning a specific weapon was the St. Petersburg Declaration of 1868, which banned explosive and incendiary projectiles weighing under 400 grams. These weapons caused suffering and injury deemed excessive to what was required to put enemy soldiers out of action. Several other conventions have followed, banning weapons as technology provided new means of killing and maiming, such as the Convention on Certain Conventional Weapons (CCW), the Chemical Weapons Convention, the Biological Weapons Convention, the Mine Ban Treaty, and the Convention on Cluster Munitions.

In the case of AWSs, the question of avoiding unnecessary suffering is arguably less applicable. Although one could imagine renegade Terminator-style robots torturing enemy soldiers or civilians, such a picture would probably be more fiction than fact.[229] I shall not speculate here on what actual weapon autonomous systems will be equipped with, but focus on the aspects unique to AWSs. The revolutionary aspect of autonomous robotics is not the weapon itself, but the operational deployment, as identification, selection, and engagement of targets would be made independently of human control. Hence, a robot using chemical weapons would clearly be illegal, but a robot using more conventional, lawful, weapons would not be illegal on those grounds.

---

[227] See K. Ipsen, 2008, pp. 78–82.
[228] API, Article 35(2). See also S. Oeter, 2008, pp. 137–9.
[229] M.N. Schmitt, 2013, p. 7; Lin *et al.*, 2009, p. 59.

4.5 The Martens Clause and Humanity

The Martens Clause is a provision included in several IHL instruments. It has its name from Fyodor Fyodorovich Martens, a Russian delegate to the Hague Peace Conferences in 1899 and 1907, and is on the face of it an attempt at filling unforeseeable loopholes in IHL. At the time of its conception, however, it was a way of finding common ground at the diplomatic conference at The Hague. The Martens Clause has featured prominently in civil society actors' case against AWSs.[230] In it, direct reference is made on the one hand to customary international law, and on the other to the somewhat more woolly 'principles of humanity' and 'dictates of public conscience'. API codifies the Clause as the following:

> In cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience.[231]

While the principles of distinction, proportionality, and unnecessary suffering have particular, well-defined meanings in IHL, the 'principle of humanity' referred to in the Martens Clause is much less precise. Although the principles of humanity appear in legal instruments, its legal basis is less strong than for the other major principles.[232] The interpretations of what the Martens Clause implies have been many. Legal scholar Antonio Cassese identifies three broad approaches:[233]

First, one might view the clause as operating on the level of *interpretation*. By this account, the invocation of 'humanity' and 'public conscience' implies that in the absence of international rules, or when doubt arises, belligerents are not free to do as they please, but are nonetheless

[230] E.g. Human Rights Watch, 2012, p. 4.
[231] API, Article 1(2).
[232] See K.M. Larsen *et al*., 2013, pp. 1–2.
[233] See A. Cassese, 2000, 189–92.

required to make humanitarian considerations.[234] While this is probably the most widespread interpretation of the Clause, such a reading reduces the Clause to a re-statement of a long-standing principle of interpretation, and is thus pointless, Cassese argues.[235] Cassese himself holds that 'humanity' should be interpreted in line with human rights law, i.e. that the 'humanity' of IHL is the same as the 'human' in human rights.[236] How one might actually apply this to IHL, however, he does not say.

Second, one might interpret the Martens Clause more radically as an independent source of international law.[237] This is apparently what HRW is doing when they contend that 'autonomous weapons would likely contravene the Martens Clause, which prohibits weapons that run counter to the "dictates of public conscience."'[238] While this interpretation seems to fit well with the wording of the Clause, there appears to be no evidence of these sources *ever* having been relied upon independently of other sources of international law either in case law or in state practice.[239]

Theorists of a third strand view the 'principles of humanity' and 'public conscience' simply as having inspired and motivated the development of IHL. This interpretation would seem to fit with the actual practice and development of humanitarian law, but it fails to explain why the Clause has been included in legal instruments. According to Larsen, Cooper, and Nystuen, the 'principles of humanity' may be more or less equated with the humanitarian considerations IHL is intended to protect. The Clause may to some extent be considered an

---

[234] Yoram Dinstein argues that rather than 'principles of humanity', the Clause ought to read 'elementary *considerations* of humanity' (italics added), as does the *Corfu Channel Case* of 1949. 'The difference between principles and considerations (or requirements) transcends semantics and goes into substance'. While principles are binding obligations, considerations point out the direction for the creation and evolution of IHL (Y. Dinstein, 2013, p. 73). By Cassese's typology, then, Dinstein subscribes to the third approach to the Martens Clause.
[235] A. Cassese, 2000, p. 192.
[236] A. Cassese, 2000, p. 207.
[237] Article 38 of the Statute of the ICJ establishes that the decisions of the Court are to be made upon the basis of (a) international conventions, (b) international customary law, (c) general principles of law, and (d) judicial decisions and legal literature as subsidiary sources of law (Statute for the International Court of Justice, Article 38). Although only formally binding for the ICJ, this definition of sources is generally considered customary for the determination of *lex lata* ('the law as it is') in the whole of public international law (A. Cassese, 2001, pp. 118–19; M. Ruud and G. Ulfstein, 2006, p. 70.).
[238] Human Rights Watch, 2012, p. 4.
[239] Y. Dinstein, 2013, pp. 72–3.

archaic remnant, as API codifies the principles the Clause alludes to in a much more precise manner.[240] The Clause would not appear to present a silver bullet for the campaigners working for a comprehensive ban on AWSs.

Nevertheless, the rules of distinction, precautions, and unnecessary suffering are normative guidelines, aiming to minimize the negative humanitarian impact of war. Some Realist commentators have argued that IHL is simply a way for states to pursue their self-interests, and that its moral content is overstated.[241] On the other hand, self-interest and morality are not necessarily opposing motivations, nor do states' intentions preclude the moral content of IHL.

What does it mean to act 'humanely'? In IHL, this injunction generally concerns the people *acted upon* in warfare; it is their 'humanity' that is due protection. The intentions or temperament – the 'humanity' – of the active side is less prominent.[242] Linking the aspect of 'humanity' to the passive side is, however, in one sense contrary to conventional speech – one is tempted to say 'public conscience' – in which one might label someone one considered to be especially cruel, numb, or without emotion as 'inhuman'.[243] Indeed, it is a thought-provoking question whether AWSs – being inanimate objects – can act 'humanely'. R.C. Arkin certainly thinks so: 'I would hope that our unmanned systems can act in a more *humane manner* and in a manner more obviously consistent with the LOW'.[244] A related question, addressed in Chapter 6 of this thesis – in which the focus is shifted from 'harm' to 'the harmer' – is whether an act can be inhumane without having inhumane effects.

---

[240] K.M. Larsen *et al*., 2013, p. 6.
[241] See K.W. Abbott, 1999, p. 365.
[242] The focus on the passive, or afflicted, side of violence appears to be a general one also in everyday language. Journalist Christopher Hitchens notes that '[o]ur common speech contains numberless verbs with which to describe the infliction of violence or cruelty or brutality on others. It only really contains one common verb that describes the effect of violence or cruelty or brutality on those who, rather than suffering from it, *inflict* it. That verb is the verb *to brutalize*. A slaveholder visits servitude on his slaves, lashes them, degrades them, exploits them, and maltreats them. In the process, he himself becomes brutalized' (C. Hitchens, 2011).
[243] As an example, consider the United Kingdom's Princess Anne's statement to the BBC, that 'gassing is the most humane way to cull badgers.' (British Broadcasting Corporation, 2014) Neither the gas nor the badger is important here, but it is interesting to note that there humane acts do not require that it is a human that is acted upon.
[244] R.C. Arkin, 2008, p. 88. Emphasis added.

4.6 Concluding Remarks on the Rules for the Conduct of Hostilities

From the discussion above, many questions remain unanswered. It appears that the current technology is far away from meeting the IHL criteria, particularly in urban conflicts and in conflicts featuring non-state actors that are unlikely to distinguish themselves from the civilian population. Of the norms discussed above, the norm of proportionality appears the most difficult, requiring AWS to be able to balance military advantage against civilian harm. However, it must be stressed that these are factual claims, susceptible to the actual technological development. The first scene on which autonomous systems could conceivably function in adherence with the principles of distinction and precautions, would probably be aerial and naval operations. According to US colonel T.X. Hammes, the typical ground operation includes 160,000 objects all involved need to be aware of, making the use of AWSs in ground operations 'extremely difficult'. In contrast, a typical air operation includes 1000 objects of concern.[245] Compared to the confusion of modern urban warfare, aerial and naval operations are more easily oriented. It is usually easier to gain oversight of the battlefield, and aircraft and ships are required to be properly marked.[246]

[245] T.X. Hammes, 2014.
[246] K. Ipsen, 2008, pp. 113–4.

PART II: THE HARMER

5 Responsibility

The technological development is currently at a stage where it is difficult to imagine how AWSs could operate in accordance with the rules for the conduct of hostilities. AWSs would presumably select their targets based on some sort of quantifiable or otherwise computer-friendly information, in that way similar to how many drone strikes are conducted today. Signature strikes and other drone strikes based on metadata and SIM card tracing are questionable considering the rules for the conduct of hostilities, as they do not appear easily squared with the principle of distinction and precautions.[247] The proportionality rule appears even more difficult for AWSs, as the balancing between ends and means required by the rule is subject to constant changes pursuant to the volatility of strategic and tactical developments. If possible at all, translating this into computer algorithms would provide a monumental task.

While the increased compliance to IHL envisioned by Arkin is endearing, it seems utopian to think that AWSs would function faultlessly. In April 2008, several TALON SWORDs – automated robots equipped with machine guns – were taken out of service in Iraq, following claims that they had turned their guns against 'friendly' soldiers. US authorities denied the story, but confirmed that there had been malfunctions.[248] In 2007, an automated cannon employed by the South African army killed nine 'friendly' soldiers and wounded 14 others.[249]

As Human Rights Watch (HRW) puts it: 'If [unlawful] killing were done by a fully autonomous weapon […] the question would become: whom to hold responsible?'[250] In an article from 2013, Aaron Johnson and Sidney Axinn have an easy answer to HRW's question:

[247] See J. Scahill and G. Greenwald, 2014.
[248] P. Lin *et al*., 2008, p. 7.
[249] P. Lin *et al*., 2008, p. 7.
[250] Human Rights Watch, 2012, p. 42.

> The chain of command is well established for all weapon systems, including remotely piloted drones. However, when the robot is controlled by a program, not a human, the usual court-martial pattern may not be useable. […] This is really a detail […]. If it is a matter of a robot mistake, for instance due to a programming error, then the same rules that govern quality assurance and verification of toasters or conventional weapons should apply. If the weapons misfired because of a design defect, the manufacturer could be held liable.[251]

On the other hand, though, we might not be content with the comparison of autonomous weapon systems to toasters, and it might be both legally and morally impermissible to try a manufacturer for war crimes. By definition, *autonomous* robots have a form of agency of their own, making the causal link between humans and the actual weapon much more blurry.

The following chapter mainly addresses the question of responsibility for war crimes in the absence of intent by humans. Possible crimes include grave breaches of the Geneva conventions, such as wilful killing, extensive destruction of property not justified by military necessity, killing or wounding a soldier *hors de combat*, and intentionally directing attacks against buildings dedicated to religion, education, art, science, or charitable purposes.[252] Trying suspects of war crimes can be very important for peace-building, norm projection, and for the justice in the eyes of the people having lost someone under terrible circumstances.[253]

5.1 State Responsibility

In traditional international law, the state as a whole bears responsibility for unlawful acts committed by all state organs, including its armed forces.[254] There is no criminal punishment in traditional international law; rulings are generally reparative, obliging the guilty state to compensate the damage done,

[251] A.M. Johnson and S. Axinn, 2013, p. 132.
[252] Rome Statute, Articles 8(a)(i), 8(a)(iv), 8(b)(vi), 8(b)(ix).
[253] See D. Luban, 2010, pp. 574–6.
[254] A. Cassese, 2001, p. 182.

and to terminate the unlawful act if it has not already been ceased.[255] This holds true also for IHL, for which Article 91 of API provides that:

> A Party to the conflict which violates the provisions of the Conventions or of this Protocol shall, if the case demands, be liable to pay compensation. It shall be responsible for all acts committed by *persons* forming part of its armed forces.[256]

In the context of AWSs, the use of the word 'persons' in API is troubling. However, the wording of the corresponding article common to the Geneva Conventions of 1949 establishes that '[n]o High Contracting Party shall be allowed to absolve itself or any other High Contracting Party in respect of breaches referred to in the preceding article.'[257] It would be reasonable to hold a state responsible for all its units in the field, regardless of them being natural persons or not.

5.2 Individual Criminal Liability?

As international law has traditionally been a matter of regulating the affairs of states, individual responsibility is historically less pronounced on the international arena. Yet, recent decades have seen an increasing individualization of responsibility also in international law through the development of international criminal law (ICL), which establishes criminal liability deriving from IHL, HRL, and domestic law.[258] ICL is less extensive than most systems of domestic law, criminal acts being restricted by the Rome Statute of the International Criminal Court to genocide, crimes against humanity, war crimes, and the crime of aggression.[259]

According to Cassese, the general principle of ICL is that no one may be held responsible for crimes committed by another person. Furthermore, a

---

[255] J. Crawford and J. Watkins, 2001, p. 285.
[256] API, Article 91. Emphasis added.
[257] Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Article 51. Also Geneva Convention (GC) II, Article 52, GC III, Article 131, GC IV, Article 148.
[258] A. Cassese, 2008, p. 6.
[259] Rome Statute, Article 5. The outlawing of the 'crime of aggression' is not yet in force.

person may be held criminally liable only if he or she is somehow culpable.[260] The Rome Statute codifies this as 'intent and knowledge'.[261] Although 'person' could imply *legal* personhood in some contexts, individual criminal responsibility in ICL should be interpreted as pertaining to *natural* persons. Indeed, the Rome Statute makes this explicit: 'The Court shall have jurisdiction over natural persons.'[262] While an AWS in theory is autonomous, and must thus be viewed as responsible for its actions in terms of strict causality, it is not a natural person, and it would seem unreasonable that an AWS could be held criminally responsible for its actions on the battlefield. Taken to its logical conclusion, such a position would require trying AWSs at The Hague or other competent courts should they be suspected of having committed international crimes.

Robert Sparrow – tongue-in-cheek – suggests that we might 'punish the culprit', or 'we might administer corporeal punishment by damaging the machine in some way, or perhaps by administering electric shocks to those electrodes through which it senses damage in combat.'[263] Assuming that AWSs would be programmed with algorithms requiring it in some way to adhere to IHL, it would furthermore be very difficult to imagine how they could breach with the law – and thus their software – with 'intent and knowledge'. Breaches would, by definition, be malfunctions.

Citing the Roman philosopher Seneca's view that '[a] sword is never a killer; it is a tool in the killer's hand',[264] Michael Schmidt asserts that the worries of HRW – that there is no one to hold responsible for the actions of AWSs – are unfounded. He identifies two channels of responsibility, that of the programmer, and that of command responsibility.

> The mere fact that a human might not be in control of a particular engagement does not mean that no human is responsible […]. A human must decide how to program the system. Self-evidently, that individual

---

[260] A. Cassese, 2008, p. 33.
[261] Rome Statute, Article 30.
[262] Rome Statute, Article 25(1).
[263] R. Sparrow, 2007, p. 72.
[264] M.N. Schmidt, 2013, p. 1.

> would be accountable for programming it to engage in actions that amounted to war crimes. Moreover, the commander or civilian supervisor of that individual would be accountable for those war crimes if he or she knew or should have known that the autonomous weapon system had been so programmed.[265]

When looking more closely at the responsibility of programmers (for international *crimes*, not commercial claims), however, the picture is complicated: Indeed, since the software is so complex, it cannot be written or micro-planned by any one person; robot software consists of millions of lines of coding written by large teams.[266] The complexity precludes anyone from having complete oversight. Hence, while the company responsible for the production of an AWS could potentially be accountable for commercial malfunction-claims, it would seems unreasonable to hold an individual programmer responsible of war crimes, unless it could be established that the malicious software was programmed with 'intent and knowledge'.[267] According to Lin *et al*., it appears very questionable whether any one programmer could be held accountable, or even have complete oversight over how the system would work in all circumstances. Even though every line of coding is internally consistent, it could prove malevolent in aggregate.

> Perhaps robot ethics has not received the attention it needs, at least in the US, given a common misconception that robots will do only what we have programmed them to do. Unfortunately, such a belief is a sorely outdated, harking back to a time when computers were simpler and their programs could be written and understood by a single person. Now, programs with millions of lines of code are written by teams of programmers, none of whom knows the entire program; hence, no individual can predict the effect of a given command with absolute certainty, since portions of large programs may interact in unexpected, untested ways.[268]

Indeed, as I have argued elsewhere in this thesis, the novelty of AWSs is that they are *not* tools, as Schmidt and others appear to hold, but must by virtue of

---

[265] M.N. Schmidt, 2013, p. 33.
[266] P. Lin *et al*., 2008, p. 8.
[267] Rome Statute, Article 30.
[268] P. Lin *et al*., 2008, p. 8.

their definition as self-governing entities be considered to have a form of agency of their own.

5.3 Command Responsibility?

The fact that AWSs themselves cannot be held individually responsible does not mean that they operate outside of the sphere of ICL. The doctrine of command responsibility, present in both IHL and ICL, ensures a hierarchical scheme of accountability, by which military and civilian leaders are responsible for the actions of the soldiers they command. This is what Schmidt has in mind when he reassures HRW, claiming that their worry is 'based on a false premise'.[269] The superior does not assume responsibility for the criminal acts of his or her subordinates directly, but may be charged guilty of having allowed a crime to happen.[270] They can, then, be held liable as aiders and abetters.[271] AP I was the first legal instrument to comprehensively codify command responsibility.

> The fact that a breach of the Conventions or of this Protocol was committed by a subordinate does not absolve his superiors from penal or disciplinary responsibility […] if they *knew*, or had information which *should have enabled them to conclude* in the circumstances at the time, that he was committing or about to commit such a breach and if they did not take all feasible measures within their power to prevent or repress the breach.[272]

As in most systems of domestic law, there are two conditions that must be fulfilled for an act to be punishable: First, a crime must have been committed (*actus reus* or 'guilty act'). This is the objective element of the crime, worded in AP I as 'a breach […] was committed'. While the objective element will not be discussed here, the second, mental, element is of more interest in our regard:

---

[269] M.N. Schmidt, 2013, p. 33.
[270] A. Cassese, 2008, p. 236. Command responsibility is thus a form of individual responsibility, not a form of substitutionary atonement.
[271] A. Cassese, 2008, p. 237.
[272] API, Article 86(2). Emphasis added.

The commander must in some way be culpable for the crime (*mens rea* or 'guilty mind'). While the Rome Statute uses similar language for the *mens rea* requirement – 'either knew or […] should have known'[273] – the wording in the Statute of the International Criminal Tribunal for the Former Yugoslavia (ICTY) is somewhat more liberal, reading as 'knew or had reason to know'.[274]

The *mens rea* for command responsibility is somewhat unclear in the legal literature and case law due to different interpretations of what 'should have known' and similar formulations actually imply.[275] It is also unclear what 'all feasible measures' implies in practice, and how to draw the line between purpose, knowledge, recklessness, negligence, and strict liability, and which of these are actually punishable under ICL.[276]

5.3.1 Singular Events

The commander is responsible for repressing crimes, and, if committed, for punishing them.[277] In the context of AWSs, punishment – at least in the conventional sense – will not be possible. The important remaining question, then, is whether civil and military commanders are in a position where they *can* in fact know whether an AWS is, or is about to, commit a crime. Puzzlingly, the *mens rea* criterion has not been much discussed in the literature on AWSs. Most authors content themselves to suggesting that there appear to be some issues regarding responsibility, or simply claim that command responsibility solves the problem.

Based on the discussion above, however, any AWS would have to be programmed with very complicated software, requiring the system to adhere to the rules for conduct of hostilities.[278] If such software could in fact be constructed, it would appear that AWSs *by definition* would have to be

---

[273] Rome Statute, Article 28(a)(i).
[274] Statute of the International Criminal Tribunal for the Former Yugoslavia (ICTY), Article 7(3).
[275] J.A. Williamson, 2008, p. 307; J.S. Martinez, 2008, p. 639.
[276] J.S. Martinez, 2007, pp. 644–5; A. Cassese, 2008, pp. 61–73.
[277] J.A. Wiliamson, 2008, p. 303.
[278] R.C. Arkin, 2010, p. 339.

expected to adhere to IHL, unless someone knew or should have known that there was a malfunction in the software. This stands in contrast to how a military commander must engage with human soldiers, as the morale and emotional state of the soldiers is likely to influence their decisions on the battlefield. This is something the commander should be expected to take into account.[279] Robots, on the other hand, would not have emotions comparable to those of humans.[280] The robot's software, would – at least in theory – restrict the system from committing war crimes. Moreover, if this software were too complicated for any one programmer to fully understand, it might be unreasonable to expect military or political commanders to do so. As the ICTY argued in *Čelebići*, 'international law cannot oblige a superior to do the impossible.'[281]

If, as suggested by Arkin,[282] so called 'case-based reasoning' is indeed programmed, a commander could conceivably be guilty of employing a too aggressive 'case', thus potentially violating the proportionality rule. For example, the actual weapon employed could be too heavy or imprecise. This would, however, negate Arkin's hope that AWSs would be incapable of violating international law through restraints in the software.[283] Anyway, if such deployment was carried through, it could potentially be viewed as 'reckless'. It is, however, not at all clear whether recklessness is included in the *mens rea* requirement in international command responsibility.[284] The suggested mental state required has varied between seemingly strict liability in *Yamashita* (*i.e.* only *actus reus* required for criminal liability),[285] to a rejection of 'criminal negligence as a basis of liability' in *Bagilishema*.[286] The recent development of ICL appears to go in the direction of a stricter interpretation

---

[279] V. Morkevicius, 2014, p. 3–4.
[280] R. Sparrow, 2007, p. 65.
[281] *Prosecutor v. Delalić et al. (Čelebići),* Judgment, 1998, Paragraph 395.
[282] R.C. Arkin, 2008, p. 64.
[283] R.C. Arkin, 2010, p. 338.
[284] I. Marchuk, 2014, p. 117–18; J.S. Martinez, 2007, p. 642.
[285] W.H. Parks, 1973, pp. 130–38.
[286] *The Prosecutor v. Ignace Bagilishema (Appeal Judgement: Reasons)*, International Criminal Tribunal for Rwanda (ICTR), 03.07.2002, Paragraphs 34–35.

than that of *Yamashita*.[287] Ordering the use of too aggressive a case could furthermore trigger a separate mode of liability called 'ordering' or 'instigating'. Constituting an act of itself, ordering or instigating would not trigger liability under command responsibility, which should be understood as a *failure* to act.[288]

In his often-cited article 'Killer Robots', Robert Sparrow argues that the only viable solution to the accountability gap is to '*assign* responsibility to an appropriate individual – presumably the commanding officer'. This would be the 'only way of meeting our obligation to enemy combatants to ensure that someone can be held responsible if they are killed unjustly.'[289] 'Assigning' responsibility in the absence of 'intent and knowledge', or some form of guilt, however, is not possible under existing ICL/IHL, and would conflict with most people's sense of justice. Substitutionary atonement is a central element of the Christian religion, but would, I contend, not be acceptable in the context of AWSs.

5.3.2 Failure to Control an AWS Over Time

As with the discussion of inherently indiscriminate weapons in Chapter 4, it could be fruitful to distinguish between singular events or immediate impact, and attacks going on over a more extended period of time. While a commander could probably not be held responsible for singular acts of an AWS (assuming that he or she did not know that there was a malfunction in the software or hardware), unlawful acts going on over a longer period of time would be more challenging. A commander who knowingly maintained an AWS with a poor track record on the battlefield over time would have to be expected to realize that there was some risk of war crimes being committed, and could then be

[287] Michael Walzer criticizes the doctrine of strict liability, claiming that when urging commanders to prevent war crimes one cannot demand success, only effort, and secondly, that the literal interpretation of the maxim that commanders should do everything they can to prevent crimes leaves them no time to do their job, which is to win wars. M. Walzer, 2006, p. 321.
[288] E.V. Sliedregt, 2011, p. 383.
[289] R. Sparrow, 2007, p. 74. Emphasis added.

held accountable. In the UK legal system, liability is determined on the basis of the *degree* of the risk. The US system, conversely, determines liability on the basis of consciousness of risk, and it is unclear which of the two should be considered *lex lata* in ICL.[290]

Elsewhere in this thesis I have argued that one of the driving forces behind the thrust for autonomy is that AWSs do not require communication and control from a headquarter. This is thought to prove beneficial in warfare against technologically capable enemies, who presumably could be able of to hack remote-controlled drones.[291] When special-operations teams operate behind enemy lines without communication with their superiors, they can at least themselves be held responsible for their acts. AWSs cannot. From the viewpoint of ICL, however, it is very questionable whether such tactics, by which a machine is left on its own without anyone even monitoring its actions, would be lawful. I suggest that it would not, as both IHL and ICL presupposes that acts of war are attributable to human beings. By extension, we may reason that any tactics that render this impossible are unlawful.

While an AWS clearly does not have criminal responsibility, either in the sense of 'answerability' or 'liability' of its own,[292] the doctrine of command responsibility also appears difficult. While it would cover some cases war crimes committed by AWSs, large loopholes would remain. Neither of the channels of responsibility identified by Schmidt appear watertight. It seems to me that only the state would be clearly responsible for the actions of an AWS. Individual responsibility both for programmers and commanders would be very hard to establish. Given the increasing importance of ICL in international law, the question then becomes whether it is lawful to develop a weapon system that to a large extent leapfrogs both individual and command responsibility.

---

[290] J.S. Martinez, 2007, p. 644.
[291] N. Sharkey, 2013.
[292] J. Crawford and J. Watkins, 2010, pp. 283–4.

6 Formulating the Uncanny: Machines with a Licence to Kill

6.1 Extending Responsibility

The previous chapter left us somewhat unsatisfied. We would have liked to have clearer lines of responsibility for the attribution of potential crimes committed by AWSs. While AWSs clearly are not human, we perceive them as something more than inanimate tools; they have some form of agency of their own. As Robert Sparrow notes, '[t]o say of an agent that they are autonomous is to say that their actions originate in them and reflect their ends.'[293] The notion of autonomy is tightly knit to the concept of *mens rea*, being subjectively aware of one's actions. In this context, Sanford Kadish has noted that '[m]uch of our commitment to democratic values, to human dignity and self-determination, to the value of the individual, turns on the pivot of a view of man as a responsible agent entitled to be praised or blamed'.[294] Allowing individual responsibility through ICL to wither, fails to acknowledge the function responsibility plays in society, and how important it could be for the people left behind to see the perpetrators (or at least those suspected) of war crimes tried.[295]

Whom or what one might include in a definition of autonomy is clearly time-dependent. One might include children, but how old? Are all grown-ups autonomous, or is a certain level of consciousness or cognitive capacity required, thus excluding people with severe enough mental disabilities? What

---

[293] R. Sparrow, 2007, p. 65. Sparrow's thoughts on autonomy and moral responsibility have kinship with David Hume's theory of agency and responsibility. According to Hume, reason and deliberation forms a *cause* of moral responsibility. Thus, 'men are not blamed for such evil actions as they perform ignorantly and casually'. Current technology has not yet evolved to feature reason and deliberation, but may in the future. D. Hume, 2004, VIII, 76. In Daniel Dennett's terminology, we may say that the precondition of moral responsibility is 'higher-order intentionality', or the capacity to form beliefs about one's beliefs. D.C. Dennett, 1997, p. 354.
[294] S. Kadish, 1987, p. 77.
[295] D. Luban, 2010, pp. 574–6.

about animals or machines? Likely, many will have experienced anger *at* (not just *with*) a computer prone to freezing at just the wrong moment. The Persian King Xerxes famously had the sea whipped for destroying the pontoon bridge he had had built over the Hellespont in the fifth century BC. A terrible storm had torn his bridge asunder, (temporarily) ruining his plans to invade Greece.[296] Almost equally foreign to our contemporary understanding are animal trials, which were practiced widely in Europe from the 13th to the 18th century.[297]

Evidently, contemporary international law fails to capture an important, perhaps *the* most important, aspect of AWSs, which is the plain objection that machines should not be delegated the capability to take moral decisions, such as selecting targets in war. With their article 'The Morality of Autonomous Robots', Aaron M. Johnson and Sidney Axinn are of the few commentators who have actually tried to justify the proposition that machines should not be allowed to kill people. They suggest that the question of decision-making is distinct from the question of responsibility, which, according to them, is 'really a detail', as AWSs are subject to the same constraints as other weapons and machines.[298] Johnson and Axinn, then, deny that AWSs have a form of agency of their own. They suggest that since 'all […] values are created by sacrifice or the risk of sacrifice', then '[w]here there is no human in the loop, there is no one to risk sacrifice, and therefore no honor produced'.[299] Moreover, a 'robot is in a way like a high-tech mousetrap; it is not a soldier with concerns about human dignity or military honor. Therefore, a human should not be killed by a machine'.[300] Lastly, '[a] nation that relies on such weapons ignores the humanitarian basis for the laws of war.' [301] This justifies a prohibition of AWSs, they contend.[302]

---

[296] Herodotus, 6.35.
[297] J.E. Salisbury, 2011, p. 115.
[298] A.M. Johnson and S. Axinn, 2013, p. 132.
[299] A.M. Johnson and S. Axinn, 2013, pp. 135, 136.
[300] A.M. Johnson and S. Axinn, 2013, p. 134.
[301] A.M. Johnson and S. Axinn, 2013, p. 138.
[302] A.M. Johnson and S. Axinn, 2013, p. 138.

Evidently, they offer three distinct arguments: (1) production of honour, (2) lack of dignity, and (3) humanitarian concerns. As I have treated the humanitarian considerations at length elsewhere in this thesis, I shall focus the former two arguments. Both of them appear lacking: First, if the production of honour (*in bello*) was the purpose of war, we should go to war as often as possible, in order to maximize our opportunities to sacrifice ourselves in the pursuit of winning praise; the injunction is prone to the same criticism as the slippery slope argument. How 'all […] values are created by sacrifice' is, moreover, not explained or justified. A sacrifice, surely, is best measured in terms of the people the sacrifice is made *for*, not the one(s) it is made *by* (which is indeed what their focus on humanitarian concerns seem to suggest). As Johnson and Axinn do not distinguish between AWSs, toasters, mousetraps, and conventional weapons on the issue of responsibility (and thus agency), it is unclear why they do so on the matter of sacrifice. By immanent logic, this would appear to tie them to the position that all weapons should be prohibited on the grounds of their missing the capacity for sacrifice. Why it is better to be saved by a sacrificing human (who might die) than a machine, also seems unclear to me, or at least worthy of justification.

A second immanent critique would be that if the proposition that all value comes from sacrifice is true, it is difficult to explain the second sub-argument, that AWSs run counter to human dignity. 'To give a programmed machine the ability to "decide" to kill a human is to abandon the concept of human dignity', they assert.[303] Through the authors' use of the word 'decide' (which is notably marked with quotation marks), and consequent decoupling of the AWS from its users, it appears that machine agency is taken back into the fold, which, surely, should have implications for their assessment of responsibility and sacrifice. Johnson and Axinn appear to want to have it both ways. (I discuss the concept of dignity later on, attempting to justify it in intersubjective recognition and moral agency.)

---

[303] A.M. Johnson and S. Axinn, 2013, p. 134.

Rather than in the capacity for sacrifice, the answer to the question of whether machines should be permitted to take human lives should be sought in the relationship between humans and between humans and machines. The objection against AWSs *themselves* (not just the harm they might inflict) goes beyond the responsibility issue strictly defined. The intermediate moral autonomy of AWSs, and the outsourcing of moral decision making their use implies, lies at the root of deeper philosophical objections to the development and deployment of AWSs. In this chapter I investigate a second main question concerning 'the harmer', eloquently put by the UN special rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns, who delivered a report on AWSs in April 2013:

> Even if it is assumed that LARs [AWSs] […] could comply with IHL, and it can be proven that on average and in the aggregate they will save lives, the question has to be asked whether it is not inherently wrong to let autonomous machines decide who and when to kill.[304]

Heyns did not himself attempt to answer this question at any length. On a meta-ethical level, one may note that what Heyns is essentially asking, is whether there are deontological or virtue ethical considerations which override the underlying utilitarian calculation.

6.2 Robots, *Automata*, and Fiction

By way of introduction, let us first take a brief look at how robots and similar agents have been looked at in the past. History – how we have reasoned about responsibility and agency in the past – might pinpoint some of the problematic issues with which we are concerned. In fiction, particularly *science* fiction, robots have had an important place. However, *Star Wars*' R2-D2 and C-3PO and the Arnold-character in the *Terminator* universe are just the culmination of millennia-old history of fictional robots or *automata*. Robot-like entities – statues coming to life or creatures being built from natural elements – are

[304] C. Heyns, 2013, p. 17.

known from both ancient Greek, Chinese, and Islamic writings. In Greek mythology, Hephaestus, the God of craftsmen, blacksmiths, and sculptors, was known to craft living creatures called *automata* for help in domestic affairs and for security. In Homer's *Iliad*, reference is made to wheel-born tripods that Hephaestus uses for domestic jobs,[305] and to golden hand-maidens, which were 'endowed with intelligence and had learned their skills from the immortal gods.'[306] In the *Odyssey*, the protagonist travels to the court of Alcinous, king of the Phaiakians, to find artificial, life-like, watchdogs guarding the palace:

> On either side stood gold and silver dogs, which Hephaestus had made with cunning skill, to keep watch over the palace of the great-hearted Alcinous and serve him as immortal sentries never doomed to age.[307]

Alcinous' dogs are clear precursors of modern weapon systems. The demand for the modern military robots is often said to be driven by the need to cover the jobs that are 'dirty, dangerous, and dull'. The watchdogs in the *Odyssey* are employed to cover the last of the 'three Ds': the dull guarding of the door. However, an important distinction should be made between Alcinous' dogs and, for instance, modern drones: While the watchdogs serve strictly *defensive* purposes, drones are, by virtue of their speed, manoeuvrability, and stamina, inherently *offensive* weapons. This allows Homer to avoid the question of the virtue of employing such robots for offensive purposes. It is fair to suggest that such tactics would not be considered all too virtuous in a tradition where soldiers went to war to win personal glory and honour (*kleos*).[308]

Thomas Hobbes famously opens his *Leviathan* with a reference to man's capacity to create an 'artificial animal', going on to ask 'may we not say, that all *Automata* […] have an artificiall life?'[309] The fascination with robots is old and carries religious undertones. Robots, *automata*, or androids are humans without a 'soul' or prospects for an afterlife.

---

[305] Homer, 18. 373–9 (2003a, p. 329).
[306] Homer, 18. 417–21 (2003a, p. 330).
[307] Homer, 7. 91–4 (2003b, p. 85).
[308] D. Hammer, 2002, pp. 210–2.
[309] T. Hobbes, 2008, p. 9.

The word robot was introduced by the Czech playwright Karel Čapek in his1921 *R.U.R.: Rossum's Universal Robots*. The play featured human-like machines, '*robota'* (meaning 'servitude' or 'drudgery' in Czech), which were put to hard labour in a dystopian world. 'Android' also has a longer history, albeit not as long as *automaton*. According to the *Oxford English Dictionary* it was first used by Ephraim Chambers' *Cyclopedia* from 1728 in reference to an *automaton*. 'Android' is composed of the Greek words for 'man' (*anēr*) and 'form' (*oeidēs)*, and in the modern usage of the term, an android is an anthropomorphic robot, *i.e.* a robot that looks like a human.[310]

The TV series *Star Trek*, first aired on the US TV channel NBC in 1966, features several so-called androids equipped with artificial intelligence, acting more or less just like any human being. Other entertainment franchises, such as the *Star Wars* films, *2001: A Space Odyssey*, and *I, Robot* followed, the latter two first appearing as book series. In many of these stories, the robots, androids, or *automata* are clearly considered to have a considerable amount of free will, responsibility, and even emotions. Although fictitious, such stories both create – and are created by – human beings' ideas and feelings about machines.

At the beginning of the twentieth century, *automata* and their likes started to become an issue not only in fiction, but also in academic literature. In 1906, the German psychiatrist Ernst Jentsch published an essay called 'On the Psychology of the Uncanny', in which he argues that *automata* and dolls have an unsurpassed aptitude in producing 'uncanny' (*unheimlich*) feelings in human beings, due to their compelling nature of begging 'intellectual uncertainty': The human observer might mistake a doll or *automaton* for a human, producing hesitancy in the observer as to how to treat, and feel about, such entities, as they (at least in our perception) assume a sort of intermediate position between being dead and alive. Jentsch focuses on E.T.A. Hoffmann's fictional story 'Der Sandmann' (The Sandman) in which the protagonist falls in

[310] OED, 2014, 'Android'.

love with a lifelike *automaton* or doll called Olympia which he only sees through a window across the street. The love-stricken young man clearly mistakes Olympia for a real woman, and the revelation and following emotional bewilderment that she is in fact not, is, according to Jentsch, the main contributing source of Hoffmann's story's uncanny atmosphere.[311]

13 years following the publication of Jentsch's essay, Sigmund Freud picked up on the same theme. Although Freud disagrees with Jentsch that the Olympia-incident is the most uncanny aspect of 'Der Sandman', Freud agrees that the uncanny (*das unheimliche*) is really a distortion of the familiar (*das heimliche*).[312] This could easily be applied to robots – not least *military* robots with lethal capabilities. Existing drones, although not physically mistakable for human beings, do display the ingredients of this mechanism: Although the actual drone is obviously made of inanimate materials, and looks nothing like a human, there is intentionality behind it. Under the drones, there is no knowing whether an operator is peering through the drone's web camera, or whether the drone is flying, and filming, on auto-pilot. Neither is there any knowing whether the drone is scanning for suspicious behaviour in preparation of a strike, and how it will react to behaviour on the ground. In one sense, the drone is just a physical extension of human intentionality, but in another sense it is an entity of its own, capable of operating for a long time without any human involvement. It seems reasonable to hold that the uncanny feelings surrounding military robots will by no means be mitigated by increased automation. Concerning AWSs, many are strongly at odds with the apparent lack of sentience in the harmer.[313]

2001: A Space Odyssey, a popular movie from 1968 telling the story of the artificially intelligent computer 'HAL', also spawned academic literature.

---

[311] S. Freud, 1990, p. 347.
[312] S. Freud, 1990, pp. 341, 371–2. Freud himself claims that the Sandman, not Olympia, is the most unsettling character of the story. Freud's interpretation, however, is rather more imaginative than that of Jentsch. According to Freud, the fear incited by the Sandman, a terrible monster who steals children's eyes and eats them, parallels the fear of castration (p. 352).
[313] P. Asaro, 2006, pp. 11–2; M. Ekelhof and M. Struyk, 2014, pp. 6–7; Human Rights Watch, 2012, pp. 36–37; R. Tonkens, 2012, 157–8; A.M. Johnson and S. Axinn, 2013, p. 134.

In a book from 1997, the philosopher Daniel Dennett rightly observed that *mens rea* 'has no requirement that the agent be capable of feeling guilt or remorse or any other emotion; so-called cold-blooded murderers are not in the slightest degree exculpate by their flat affective state.'[314] According to Dennett, the requirement for moral responsibility is 'higher order intentionality', meaning the capability of framing beliefs about one's own beliefs.[315] While no computer software has that ability today, they may get it in the future. In that case, Dennett argues, it is difficult to see how machine agency differs from that of humans.[316] John Sullins goes even further than Dennett, submitting that when a robot has 'autonomous intentions and responsibilities […], then the machine is a robust moral agent, possibly approaching or *exceeding the moral status of human beings.*'[317] On the one hand, we may on a rational level be inclined to agree with the conclusions of Dennett and Sullins, but on the other hand, we might still feel that their inanimate nature precludes them from full membership in the 'moral sphere', which I, following Aristotle, define as the social institution in which praise and blame is distributed.[318]

6.3 Towards An Intersubjective Theory of Killing

Although the members of the Campaign to Stop Killer Robots have largely tried to frame the matter as one of potentially adverse humanitarian consequences,[319] I suggest that it is the 'uncanny' that is the basis the opposition to AWSs. As experts on robotics are in disagreement, and AWSs are not yet in existence, we simply do not know what the effects of their deployment will be. (While for example anti-personnel mines and cluster munitions by definition cannot be used in accordance with the principle of distinction, no such definitional constraint exists for AWSs.) If they could in

[314] D.C. Dennett, 1997, pp. 351–2.
[315] D.C. Dennett, 1997, pp. 354.
[316] D.C. Dennett, 1998, pp. 154–9.
[317] J.P. Sullins, 2006, p. 29. Emphasis added.
[318] Aristotle, 2002, III.1. See also R. Sokolowski, 2001, pp. 355–7.
[319] Human Rights Watch, 2012; M. Ekelhof and M Struyk, 2014; N. Sharkey 2007, 2008.

fact not be used in a discriminating manner, they would already be considered unlawful under international law (see Chapter 4). Notwithstanding this, many still feel unrest and even 'feelings of revulsion'[320] at AWSs taking human lives. A survey of US public opinion from 2013 found that, while 18 per cent of the respondents were undecided, 55 per cent strongly or somewhat opposed the use of autonomous weapons.[321]

As we saw in Chapter 4, notions of 'humanity' are strongly present in IHL, particularly through the Martens Clause. While the 'humanity' or 'inhumanity' referred to in the Clause in *legal* terms should not be seen in isolation from negative humanitarian consequences, an alternative reading of the clause, whereby focus is shifted from actions and consequences to the actors themselves, might be necessary to understand the curious case of AWSs.[322] Perhaps killing by a non-human is in a way inhumane as such?

Codes of chivalry and honour – essentially virtue ethical ideals – have set the course for the distribution of praise and blame in war for centuries.[323] Such virtue ethical ideals are highly relational; they are impossible to see in isolation from the context in which they are practiced, and towards whom a given act is directed.[324] In a 1972 paper, Thomas Nagel asserted that there is a certain intersubjective, or interpersonal, dimension to the ethics of war:

> [H]ostility or aggression should be directed at its true object […] It is evident that some idea of the relation in which one should stand to other people underlies this principle, but the idea is difficult to state. I believe it is roughly this: whatever one does to another person intentionally *must be aimed at him as a subject, with the intention that he receive it as a subject*.[325]

---

[320] M. Ekelhof and M Struyk, 2014, p. 6.
[321] C. Carpenter, 2013.
[322] See Chapter 4 in this thesis and e.g. K.M. Larsen *et al*., 2013, p. 6 for a discussion of the legal interpretation of the 'principle of humanity'.
[323] Chivalry, courage, and wisdom were essential features of ancient Greek philosophy. E.g. Aristotle, 2002, III,1.
[324] G.E.M. Anscombe, 2007, pp. 26–30.
[325] T. Nagel, 1972, pp. 135–36. Emphasis added.

What does it mean that the harm must be directed at a subject? Nagel himself does not say. Accordingly, the following sections is an attempt at giving flesh to the bone to the notion that even warfare should be viewed as a relation between subjects, and that a form of warfare that distorts such a relation is unethical.

6.4 The Master–Slave Dialectic

Nagel's take on aggression displays kinship with the Hegelian conception of self-consciousness as developed through a 'struggle for recognition'.[326] Hegel's master–slave dialectic offers a powerful heuristic for our understanding of the uncanniness of AWSs.[327] Before diving into Hegel's philosophy, however, it is important to note that the master–slave dialectic (and the contemporary theory of recognition, which builds on Hegelian philosophy),[328] has two sides to it: On the one hand, it is an analysis of *empirical*, actual, patterns of human relations in the world. On the other hand, it has powerful *normative* implications.[329] In other words, Hegel tries to show how his ethics are rested on the material world as well as in idealist thought (the latter being what he is most famous for).

According to Hegel, the world is structured around contradictions (both in terms of how it unfolds and how we grasp it epistemologically).[330] A thing is never just something in *itself*; it assumes objective existence only in relation to its negation.[331] This also holds true for individual human beings and human societies, as they seek recognition of their humanity and existence from others;[332] as Hegel puts it in the Jena lectures, only when recognized does the individual have existence. Humans demand an acknowledgement of their

---

[326] A. Honneth, 1995, p. 5; G.W.F. Hegel, 1952, pp. 140–3.
[327] Hegel's discussion of the formation of self-consciousness, and its distortion in the master and slave relation is found in Hegel's early Jena lectures (1805–06), and in *Phänomenologie des Geistes* (1807).
[328] See C. Taylor, 1994; 2005, Chapter XX; A. Honneth, 1995; 2003; 2010; 2011; McNay, 2008.
[329] L. McNay, 2008, p. 8. Following conventional terminology, is thus a *social* theory more than strictly a *political* theory. The lines drawn between these disciplinary branches, are, however, often drawn 'very much in the sand'. M. Philip, 2010, p. 129. See also L. McNay, 2010, pp. 85–7.
[330] C. Taylor, 2005, p. 226.
[331] G.W.F. Hegel, 1963I, pp. 30–31.
[332] G.W.F. Hegel, 1952, p. 140.

identities as something more than just natural existence; a recognition that they are, as it were, not immanently 'attached to life'.[333] As Hegel sees it, however, successful mutual recognition, by which both subjects are recognized as worthy, self-containing individuals, is not always possible. In fact, individuals will typically try to force recognition from the 'other'. As Hegel puts it in the *Phenomenology*, 'the relation of the two self-conscious individuals is such that they prove themselves and each other through a life-and-death struggle'.[334] Particularly in Hegel's early writings this mechanism of a struggle for recognition is held to count also for collectives such as societies, communities, and sub-cultures.[335]

In the life-and-death struggle for recognition, two outcomes are possible: First, one or both subjects are killed, or second, one of them capitulates, thus becoming the other's slave. 'Through successfully risking his life, the master proves that he transcends the natural world, and his own natural existence.'[336] In Hegel's thought, this is a tragedy, as the relation of lordship and bondage precludes both the slave from being recognized (which would undo the master–slave relation) but also the master from true recognition from a peer. Recognition from a slave becomes worthless, as the slave is reduced to an object of natural existence.[337] Hence, both the master and the slave remain unfree, and they both fail in developing their identities and integrity. The master is furthermore never safe from being ambushed by the slave, and is alienated from the external world through reliance on the slave.[338] Thus, in the typical Hegelian vein, there occurs a reversal; 'mastery, as sheer self-affirmation, turns out to be self-subverting and brings about the opposite of what it intended. Mastery ends in failure, a dead end that can only be maintained by force.'[339]

---

[333] G.W.F. Hegel, 1977, p. 113.
[334] G.W.F. Hegel, 1977, pp. 113–4.
[335] A. Honneth, 1995, p. 5.
[336] R.R. Williams, 1997, p. 62.
[337] G.W.F. Hegel, 1977, p. 116; R.R. Williams, 1997, p. 63.
[338] G.W.F. Hegel, 1969, pp. 212–6; 1952, pp. 140–43; C. Taylor, 2005, p. 156; A. Kojève, 1980, p 30.
[339] R.R. Williams, 1997, p. 64.

Hegel's dialectic of lordship and bondage has been subject to very different interpretations. While Alexandre Kojève interprets all human relations to at some level reflecting master and slave relations,[340] Robert Williams reads the passage as a prudential account of failed recognition:[341] 'The fundamental inequality of the substance-accident relationship is institutionalized in the intersubjective social shape of domination and submission, lordship and bondage'.[342] Charles Taylor, to the contrary, interprets the Dialectic as an account of a historical period corresponding to the Hobbesian state of nature.[343]

Anyway, Hegel's intersubjective account of the master and slave relation provides us with an interesting analogy for the case of AWSs, which may enable us to further understand the uncanniness of AWSs. For simplicity, I shall in the following discuss a situation in which only one of the parties to the conflict employs AWSs.[344]

Let us imagine that a conflict between two parties arises. Imagine that one of the parties (e.g. 'South') is utilizing AWSs and the other is not (e.g. 'North').[345] Imagine that through technological superiority and the use of AWSs, South defeats North and subverts it. South further demands compliance and submission from North. Notwithstanding its technological edge, however, South, the master, remains vulnerable to ambush and indiscriminate attacks from the subjected North, the slave.

The argument that the use of AWSs is likely to redistribute harm from soldiers to civilians has been made by several contributors in the field,[346] and was partly dismissed by this author in Chapter 3. Yet while the argument here is similar, it is not identical. As part of the broader slippery slope argument, the asymmetry-problem was viewed through the lens of *consequences*; here it is viewed independently from consequences, through the lens of *agency*. There is

---

[340] A. Kojève, 1980, pp. 12–17.
[341] R.R. Williams, 1997, p. 63.
[342] R.R. Williams, 1997, p. 62.
[343] C. Taylor, 2005, p. 153. See T. Hobbes, 2008, Chapter XIII.
[344] This could be viewed as constituting an ideal assumption. See Chapter 2.
[345] We may indeed also imagine that both parties are using AWSs. The point is that AWSs are being employed against humans.
[346] J.C. Galliott 2010a, 2010b, C.G. Kels, 2010; A.M. Johnson and S. Axinn, 2013.

something in the relation between humans and machines itself that appears unsettling. Here lies a key to our understanding of AWSs.

Through the use of AWSs, South is blocked from the moral sphere and intersubjective relations. To South, war has become like an industrial activity. There is no emotion, no virtue. Perhaps, though, a sense of shame is forming. As Nobel laureate Jody Williams has admitted: 'When my country wants to call it a bloodless battlefield I feel enraged. I feel righteous indignation at the twisting words.'[347] Both parties fail in recognizing each other because of the unbridgeable gap created by technology.

As in Hegel's thought, we see the outline of a situation whereby the master is alienated from the external world as she refuses to partake in the actual fighting, rather outsourcing it to the slave and the machine. Here lies an important ambiguity and difference: In Hegel's Dialectic, the master gains supremacy because she does not fear death: This is how she seeks to transcend natural existence in order to have his independence, identity, and integrity recognized. When using AWSs, however, the opposite is true. The party employing AWSs is in essence fleeing the battlefield precisely because it – or its political leaders – fear death. Thus, the master in the South–North analogy may not be the master at all; an even clearer reversal than in the *Phenomenology* is evident. '[i]t is only through staking one's one life that freedom is won.'[348]

In Albert Camus' play *The Just Assassins* (1949) the issue of risk is explored: The protagonists, Kaliayev, purports that even an assassination can be justified as long as the assassin exposes him- or herself to risk, and faces up

[347] Quoted in M. Ekelhof and M. Struyk, 2014, p. 6. An ideal type of robotic warfare that doesn't affect human beings seems wildly unrealistic, and quite frankly pointless, as the whole purpose of war is to hurt people in order to bring some outcome. Carl von Clausewitz exposed this almost 200 years ago: 'Kind-hearted people might of course think there was some ingenious way to disarm or defeat the enemy without too much bloodshed […]. Pleasant as it sounds, this is a fallacy that must be exposed: war is such a dangerous business that the mistakes which come from kindness are the worst. […] This is how the matter must be seen. It would be futile – even wrong – to try and shut one's eye to what war really is.' C. Clausewitz, 1984, pp. 75–6.
[348] G.W.F. Hegel, 1977, p. 114.

to the consequences after the deed is done.[349] More than just physical risk, taking a life requires moral responsibility. The CIA's use of remote-controlled drones would be a good example of this. When the United States use drones outside of war zones it is not the US military that uses them, but the CIA, which is not subject to the same reporting obligations as the military. The use of AWSs, however, goes a step further than leapfrogging physical risk (like artillery) and responsibility (like CIA-controlled drones): The use of AWSs lead to an outsourcing of moral agency to machines. Indeed, we might go as far as to say that AWSs are not riskless weapons at all, as such a concept at least to a certain extent requires that *someone* is relieved of risk. In the case of remote-controlled drones it is obvious who this someone is (the pilot), but in the case of AWSs he or she is just an abstract unknown, someone who might have performed the same tasks as the AWS in the past.

The CIA's (and to a somewhat lesser extent the Pentagon's) use of drones has been the subject of stark opposition from international organizations including the UN, the EU, a number of civil society actors, and media,[350] using phrases such as 'inherently un-American and cowardly', and 'a danger to humanity'.[351] Much of this criticism has come from within the United States and other countries that employ drones, so it is clear that there is a sense of embarrassment on one's own state and people's behalf. I argue that the use of AWSs would take this even further. As Hegel purports in the master–slave dialectic, the tragedy is not just that the slave is misrecognized, but also that the master is barred from the moral sphere. The same is happening when moral agency is outsourced to machines. I argue that the distortion of the master–slave dialectic evident in the hypothetical use of AWSs contribute strongly to the feelings of 'revulsion' at automated killing described by some commentators. The important point here is that *the use of AWSs may not only*

[349] A. Camus, 1958, Act 3. See also the author's preface and J. Walter, 1989, p. 568.
[350] E.g. P. Alston, 2010; European Parliament Resolution on the Use of Armed Drones, 2014; Reprieve, 2014; M.A. Shah, 2014; G. Miller, 2014.
[351] A. Iqbal, 2012; G. Monbiot, 2012.

be a moral wrong only against the people against whom they are used, but also against the people using them (and the people on whose behalf they are used).

In that process humans, as it were, become machines, and *vice versa*. While machines have commonly been thought of as aids for human beings – technologies to facilitate human agency – the use of AWSs turns that relationship on its head: Humans would presumably do the facilitating – programming, coordinating, washing, fuelling – and machines would be taking the moral decisions. In his essay 'Shooting an Elephant', George Orwell ponders the (Hegelian) mechanism by which actions reflect back upon the actor in unexpected ways. Writing about his experiences in Lower Burma, Orwell explains his feeling of misrecognizing himself:

> Here was I, the white man with his gun, standing in front of the unarmed native crowd – seemingly the leading actor of the piece; but in reality I was only an absurd puppet pushed to and fro by the will of those yellow faces behind. I perceived in this moment that when the white man turns tyrant it is his own freedom that he destroys. He becomes a sort of hollow, posing dummy, the conventionalized figure of a sahib.[352]

From Orwell's point of view, it is not always obvious who is the oppressor and who is the oppressed.

6.5 Arendt's Active Life

The German–American philosopher Hannah Arendt (influenced by the Hegel's concepts of action and consciousness)[353] similarly understands human relations as intersubjective. In *The Human Condition* (1958) Arendt attempts to conceive of what it is that makes humans human. According to her, human activities can be divided into three: (1) labour, (2) work, and (3) action. What truly distinguishes humans from other species is the latter category of 'agency'. While labour comes down to the rudimentary maintenance of physical existence, work is the manufacturing of things, the creation of a separate

[352] G. Orwell, 1981, p. 152. *Sahib* is Arabic, meaning, as it happens, 'master'.
[353] A. Speight, 2002.

human world.[354] Action, however, transcends the instrumental nature of the two former categories, as it is 'an end in itself'.[355] Arendt argues that it is a mistake to take freedom to be primarily an inner, contemplative or private phenomenon, for it is in fact active, worldly and public'.[356] Action can only be understood as a joint enterprise, as the transcendence of the instrumental dimension necessitates a society or community. It is intrinsically linked to what it means to be human, and what it means to be free.[357]

Already in the mid-twentieth century, however, Arendt saw a grave challenge to ethical life in the form of 'the advent of automation'.[358] While automation can certainly have its values in industry, assuming roles in labour and work, the sphere of action is inherently political, and inherently human, and accordingly, should not be left to machines.[359] Abandoning the sphere of action could on the one hand have negative moral consequences, but more importantly – and in line with what we discovered in the last section – would also imply the disempowerment and alienation of the humans responsible for the withdrawal from the world. In other words, allowing machines to take over human agency is to disrespect humanity *in its most basic integrity and nature* as moral and political agents. Put differently, the use of AWSs is a fundamental negation of our own nature. This is precisely what the use of AWSs would imply; it would constitute a retreat from the moral world, amounting to declaring oneself unfit to manage one's own affairs. It could also, in existentialist terms, be labelled *cowardly*, if decision-making is seen with Sartre (as Arendt) as a fundamental aspect of ethical life.[360] Again, the objection is not to the lack of *physical* risk, but the lack of *moral* risk.

In *Eichmann in Jerusalem* (1963), Arendt's account of the trial of one of the main architects behind the *holocaust*, she claims that the 'trouble with

---

[354] Arendt is here influenced by Martin Heidegger, who distinguished between the (natural) 'earth' and the human (created) 'world'. A.J. Vetlesen, 1996, p. 14.
[355] H. Arendt, 1973, p. 153.
[356] M. Yar, 2013.
[357] H. Arendt, 1973, pp. 188–90; A. Speight, 2002, p. 531.
[358] H. Arendt, 1973, p. 4.
[359] H. Arendt, 1973, pp. 175–6.
[360] J.P. Sartre, 2010, pp. 67, 574; H. Arendt, 1978, p. 180.

Eichmann was precisely that so many were like him, and that the many were neither perverted nor sadistic, that they were, and still are, terribly and terrifyingly normal'.[361] According to his own moral compass, Eichmann was not guilty of anything; he had fled the sphere of moral decision-making, taking refuge in orders and duty. 'The longer one listened to him [Eichmann], the more obvious it became that his inability to speak was closely connected with an inability to think.'[362] This is, according to Arendt, the 'banality of evil'; it arises from a failure to make moral choices, in the sense of withdrawing from, or narrowing, the scope of moral action. That was what made the *holocaust* so evil. 'The sad truth of the matter is that most evil is done by people who never made up their minds to be or do either bad or good'.[363]

Hannah Arendt identified two types of 'modern world alienation', 'twofold flight from the earth into the universe and from the world into the self'.[364] Firstly, humans have tried to extricate themselves from the external world through automation, and secondly, they have escaped into themselves, isolating themselves from the moral and political world.[365] The use of AWSs unites this 'twofold flight' into a single retreating motion.

6.6 Recognition and Dignity: Beyond Lordship and Bondage

According to Hegel, 'ethical life' and freedom is represented by three institutions: (1) relations of love in within the family, (2) contractual respect in the civil society, and (3) notions of solidarity in the state.[366]

According to the German social theorist Axel Honneth, all normative relations can be traced back to, and modelled as, questions of recognition.[367] Expanding on the philosophy of Hegel and the social psychology of George H.

---

[361] H. Arendt, 2006, p. 276.
[362] H. Arendt, 2006, p. 49.
[363] H. Arendt, 1978, p. 180.
[364] H. Arendt, 1973, p. 6.
[365] H. Arendt, 1973, pp. 253–5.
[366] G.W.F. Hegel, 1902, §§ 37, 183, 260.
[367] A. Honneth, 2003, pp. 170–4.

Mead, Honneth launches a moral theory built on an analytic foundation.[368] Honneth subscribes to Hegel's account of the development of identity and ethical life in the three stages, and argues that they have the distinct functions of producing self-confidence, self-respect, and self-esteem.[369] Herein lies also the basis of disrespect, Honneth postulates: For each positive source of mutual recognition and positive relation-to-self – love, contractual or legal respect, and solidarity – a corresponding source of misrecognition can be identified: (1) physical maltreatment, (2) denial of rights or social ostracism, and (3) downgrading of social and moral value.[370] As we saw in the section on the master–slave dialectic, social relations are reciprocal, meaning that denial of recognition to someone else may inhibit our own moral integrity and identity.

Due to the denial of mutual recognition and retreat from the moral sphere it constitutes, the use of AWSs may be said to represent a disrespect of the enemy, first, on the level of physical maltreatment, and second, on the level of social value and solidarity. Being physically maltreated by an agent with whom moral relations are impossible constitutes an ethical challenge beyond 'normal' killing, because of the lack of sentience and possibility of reasoning with one's opponent. The complete withdrawal from the moral sphere pushes war in the direction of 'vermin control', by which a problem is exterminated automatically, asymmetrically, and without feelings. This, I argue, is likely to produce an extreme sense of disrespect and humiliation, the root of injustice in the theory of recognition.[371]

On the level of social value, the use of AWSs may be said to undermine the dignity of the party against whom they are used. It may be thought casuistry to distinguish whether someone is killed by a machine or by a human – the fundamental issue at stake, surely, is the killing – but on the other hand, there is a long tradition of honour and dignity as virtues in warfare, probably more so in warfare than in domestic law enforcement (see footnote 145). The phrasing

[368] A. Honneth, 1992, p. 192; L. McNay, 2010, pp. 85–6.
[369] A. Honneth, 1995, Chapter 5.
[370] A. Honneth, 1992, p. 193–5.
[371] A. Honneth, 1992, p. 189; C. Heyns, 2013, p. 17–8.

of the first sentence in the Universal Declaration of Human Rights powerfully captures the essence of the underlying principle, expressing a 'recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world'.[372]

The issue of recognition spills into the question of responsibility: Offering recognition, or at least the possibility of an intersubjective relation, is at the same time to be morally responsible. Acts of war are inherently moral acts, to which there should be attached clear lines of responsibility. As Michael Walzer puts it: 'There can be no justice in war if there are not, ultimately, responsible men and women.'[373]

Continuing with the just war tradition, one may furthermore say that that the use of AWSs gravely challenges the notion of the moral equality of soldiers, a central concern in the modern just war tradition following Michael Walzer. According to him, individual soldiers have a moral obligation to act in consistence with the law and moral codes *on* the battlefield (*in bello*), but are exempt from the justness or unjustness of the war as a whole (*ad bellum*). The fundamental distinction in warfare is not between the guilty and the innocent, but between combatants and non-combatants. The licence to kill innocents (which regular soldiers most often are) is usually theorized as legitimized by the mutual imposition of risk,[374] but, I argue, this licence also requires 'moral risk', *i.e.* responsible agency.

The laws and morality of war presuppose a reciprocity and moral equality between the belligerents. This equality is eroded by riskless warfare. The use of armed drones, then, may be said to be *im*moral. The use of AWSs, however, goes beyond this injustice: Withdrawing from the moral sphere through the outsourcing of moral decision-making *a*moralizes war, making the question of moral equality irrelevant or inappropriate. According to both the

---

[372] Universal Declaration of Human Rights, 10 December 1948, Preamble. The Universal Declaration is not legally binding, but is commonly held to codify customary law. See J. Rehman, 2010, pp. 80–1.
[373] M. Walzer, 2006, p. 288.
[374] P.W. Kahn, 2002, p. 2.

Hegelian theory of recognition and the just war tradition, morality cannot be understood other than intersubjectively. The moral equality of soldiers has what we might see as a corresponding Article in the Universal Declaration of Human Rights, which again, seems difficult to square with the use of AWSs against humans: 'All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.'[375]

If the principles of reciprocity, recognition, and dignity are indeed as absolute as theorists like Axel Honneth claim, they should not be put at risk. The use of AWSs would be unethical, since such weapons would bar humans from the 'Arendtian' sphere of action, and would constitute a misrecognition of the physical integrity and social worth not only of the enemy, but also of the people using them, and the people on whose behalf they are applied.

[375] Universal Declaration of Human Rights, 1948, Article 1.

7 Conclusion

The questions pertaining to whether AWSs would adhere to the rules for the conduct of hostilities on the one hand, and the supposed consequences of their deployment on the other, are qualitatively different from the question of whether machines ought to be granted the capability to take life-or-death decisions. In this thesis, I have tried to separate these questions.

While the former two questions belong within a consequentialist universe (as they implicitly answer the latter question affirmatively), the latter belong to the world of duties and virtue. While the former questions are susceptible to changes in the external world, most importantly the technological development of artificial intelligence, the answer to the latter question *cannot* be resolved by anything in the empirical world. While the former questions can be, and are, opposed on factual grounds, the latter question begs a more definite answer.

In the chapter on the slippery slope – the notion that allowing the use of AWSs will lead to a host of unintended, negative consequences such as a lower threshold to resort to the use of force in international affairs – I noted that there are crippling objections against most of these arguments. First, the argument that AWSs increases the inclination to use force has difficulties distinguishing between AWSs and other so-called riskless weapons, or indeed other technologies such as armour, camouflage, and medicine, which are all intended to protect soldiers. Perhaps the use of both AWSs and medicine in war *is* morally wrong, but it seems to me that a theory addressing this would be an ideal one for an ideal universe.[376] Second, while the argument that AWSs may redistribute harm from soldiers to civilians could be valid, hypotheticals seldom do enough to justify the prohibition of a weapon deemed to have military utility.

[376] See A. Swift and S. White, 2010, p. 58.

In the chapter on the conduct of hostilities, I noted that given the definition of AWSs as 'weapons that can independently select and attack targets', the important question is not whether such systems would be inherently indiscriminate (by definition, they would be, as only military objectives and soldiers may be defined as 'targets'[377]), but whether it is possible to develop such machines at all.

I noted that current technology is far away from such a prospect. The norm of distinction requires that parties 'shall direct their operations only against military objectives.'[378] In the context of AWSs, a robot or software would be the one 'directing'. As targeting is a fundamentally strategic act, AWSs would to some extent be granted power over the dynamic and development of the war itself. Particularly the norm of proportionality appears very difficult, considering that the system itself would have to be able to balance fluctuating interests and unintended civilian casualties. Nevertheless, the argument is an empirical one, and improved technology would obligate its subscriber to allow for their use. While for example anti-personnel landmines are discriminate by definition, AWSs are not.

In the chapter on responsibility I shifted focus from the consequences of the use of AWSs – the harm they might inflict – to the objections against the weapon systems themselves: 'the harmer'. Here, I contended that AWSs to a considerable degree pulverize individual responsibility under IHL/ICL. Indeed, it is possible to imagine a situation whereby an AWS by all intents and purposes 'intentionally' commits a war crime without any one person responsible. The software an AWS requires would be so complex that it would be physically impossible for an individual to read all the lines of coding.[379] On the other hand, individual responsibility is not a *vital* part of international law. State responsibility, which AWSs do not appear to challenge, would still apply.

---

[377] S. Casey-Maslen, 2014, forthcoming.
[378] API, Article 48.
[379] P. Lin *et al*., 2008, p. 8.

In the last chapter, I looked at the objection against giving AWSs a 'licence to kill', the argument that it is inherently wrong to allow machines to kill people. I claimed that such an argument ought not to be decoupled from the question of responsibility, nor justified with AWSs' lack of a capacity for sacrifice.[380] Rather, a prohibition of AWSs should be justified in the retreat from the moral sphere that their use implies. The use of AWSs would arguably be a moral wrong not just against the people they are used, but also against their employers. It would contradict and misrecognize their status as moral agents, and alienate them from the moral significance of the war. The application of AWSs would imply an outsourcing of 'agency', according to Arendt the defining aspect of humanity, consequently running the risk of the 'banality of evil'. Their use would furthermore impair the moral equality of soldiers, and the dignity of both the user and the victim. I suggested that the intermediate status of AWSs between being dead and alive, and the distortions of intersubjectivity referred to above, produce the uncanny feeling that I believe is the major driving force behind the mounting opposition to AWSs. Hence, I suggested that the last of the arguments is the strongest one in the quest for a prohibition of AWSs.

Whether 'the harmer' can be decoupled from 'the harm' he or she causes is the primary question that states and their leaders must ponder as the grounds for a ban on AWSs are discussed. I would suggest that while IHL does not necessarily fail to regulate AWSs, it might still be insufficient in capturing the central challenge posed by machines with a licence to kill. This is the fundamental question both individuals and states must ask themselves, lest warfare might be changed irreversibly for the worse. AWSs raise urgent questions about moral agency that the international community must address.

[380] A.M. Johnson and S. Axinn, 2013, pp. 135–6.

8 Literature

Abbott, K.W., 'International Relations Theory, International Law, and the Regime Governing Atrocities in Internal Conflicts', *American Journal of International Law*, vol. 93, no. 2, 1999, pp. 361–79.

Abney, K., 'Autonomous Robots and the Future of Just War Theory' in F. Allhoff, N.G. Evans, and A. Henschke (eds)*, Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century*, London, Routledge, 2013.

Air–Sea Battle Office (US Department of Defense), 'Air–Sea Battle: Service Collaboration to Address Anti-Access & Area Denial Challenges', 2013, http://navylive.dodlive.mil/2013/06/03/overview-of-the-air-sea-battle-concept/, (accessed 31.01.2014).

Alston, P., 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions: Study on Targeted Killings, Geneva, United Nations, 2010.

Annex to Convention (IV) Respecting the Laws and Customs of War on Land: Regulations Concerning the Laws and Customs of War on Land, the Hague (enacted 18.10.1907, in force 26.01.1910).

Anscombe, G.E.M., 'Modern Moral Philosophy' in R. Crisp and M. Slote (eds), *Virtue Ethics*, Oxford University Press, 2007, pp. 26–44.

Arendt, H., *The Human Condition*, University of Chicago Press, 1973 [1958].

Arendt, H., *The Life of the Mind*: *Thinking*, M. McCarthy (ed), Boston, MA, Houghton Mifflin Harcourt, 1978.

Arendt, H., *Eichmann in Jerusalem: A Report on the Banality of Evil*, London, Penguin, 2006 [1963].

Aristotle, *Nicomachean Ethics*, trans. C. Rowe, Oxford University Press, 2002 [ca 350 BC].

Arkin, R.C., 'Governing Lethality: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture', Technical Report GIT-GVU-07-11, *U.S. Army Research Office*, 2008, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6249424&tag=1, (accessed 15.08.2013).

Arkin, R.C., 'The Case for Ethical Autonomy in Unmanned Systems', *Journal of Military Ethics*, vol. 9, no. 4, 2010, pp. 332–41.

Asaro, P., 'What Should We Want from a Robot Ethic?', *International Review of Information Ethics*, vol. 6, 2006, pp. 9–16.

Asaro, P., 'On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization or Lethal Decision-Making', *International Review of the Red Cross*, vol. 94, no. 886, 2012, pp. 687–709.

Auden, W.H., 'August 1968' in E. Mendelson (ed.), *W.H. Auden: Selected Poems*, New York, Vintage, 2007.

Augustine, *City of God*, trans. Henry Bettenson, London, Penguin, 2003 [426].

Becker, J., and S. Shane, 'Secret "Kill List" Proves a Test of Obama's Principles and Will', *The New York Times*, 29.05.2012, http://www.nytimes.com/2012/05/29/world/obamas-leadership-in-war-on-al-qaeda.html?pagewanted=all&_r=0, (accessed 12.09.2013).

Begby, E., G. Reichberg, and H. Syse, 'The Ethics of War. Part I: Historical Trends', *Philosophy Compass*, vol. 7, no. 5, 2012, pp. 316–27.

Bekey, G.A., *Autonomous Robots: From Biological Inspiration to Implementation and Control*, Cambridge, MA, MIT Press, 2005.

Besson, S., 'Theorizing the Sources of International Law' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 163–85.

Besson, S. and J. Tasioulas, 'Introduction' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 1–27.

BBC (British Broadcasting Corporation), 'Princess Anne: Gassing Badgers is the Most Humane Way to Cull', 05.04.2014, http://www.com/news/uk-26888406, (accessed 09.05.2014).
Brehm, M., 'CCW Report: The Need for New International Law', http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2014/ccwreport/CCWR1.2.pdf, (accessed 14.05.2014).
Brooks, M., 'Meet the Man Who Wants Total Unemployment for All Human Beings in the World', *The New Statesman*, 19.09.2013, (accessed http://www.newstatesman.com/economics/2013/09/meet-man-who-wants-total-unemployment-all-human-beings-world).
Bureau of Investigative Journalism, 2014, http://www.thebureauinvestigates.com/category/projects/drones/, (accessed 27.03.2014).
Campaign to Stop Killer Robots, 'Learn', 2013, http://www.stopkillerrobots.org/learn/, (accessed 14.05.2014).
Camus, A., *Caligula & Three Other Plays*, trans. S. Gilbert, New York, Vintage Books, 1958.
Canning, J., 'A Definitive Work on Factors Impacting the Arming of Unmanned Vehicles', *Dahlgren Division US Naval Surface Warfare Center*, Report NSWCDD/TR-05/36.
Carpenter, C., 'US Public Opinion on Autonomous Weapons', *University of Massachusetts*, *Department of Political Science*, survey conducted by YouGov America, 10/12.05.2013, http://www.whiteoliphaunt.com/duckofminerva/wp-content/uploads/2013/06/UMass-Survey_Public-Opinion-on-Autonomous-Weapons.pdf, (accessed 14.03.2014).
Casey-Maslen, S., 'The Use of Nuclear Weapons Under Rules Governing the Conduct of Hostilities' in G. Nystuen, S. Casey-Maslen, and A.G. Bersagel (eds), *Nuclear Weapons Under International Law*, Cambridge University Press, forthcoming 2014.
Cassese, A., 'The Martens Clause: Half a Loaf or Simply Pie in the Sky?', *European Journal of International Law*, vol. 11, no. 1, 2000, pp. 187–216.
Cassese, A., *International Law*, Oxford University Press, 2001.
Cassese, A., *International Criminal Law*, Oxford University Press, 2008.
Chambers, R., 'The War of the Words: The Rhetoric of "Operation Iraqi Freedom" (an informal survey)', *Culture, Theory and Critique*, vol. 44, no. 2, 2003, pp. 171–81.
Charter of the United Nations (enacted 26.06.1945, in force 24.10.1945).
Clausewitz, C. von, *On War*, M. Howard and P. Paret (trans and eds), Princeton University Press, 1984 [1832].
Cobb, T., 'Good-Bye Counter-Insurgency; Hello Air–Sea Battle', *Harvard National Security Journal*, 2011, online feature, Journal, http://harvardnsj.org/category/online/features/page/4/, (accessed 08.04.2014).
Coeckelbergh, M., 'Drones Information Technology, and Distance: Mapping the Moral Epistemology of Remote Fighting', *Ethics and Information Technology*, vol. 15, no. 2, 2013, pp. 87–98.
Constitutional Rights Foundation, 'Firestorms: the Bombing of Civilians', 2013.
Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, Paris 13 January 1993 (enacted 13.01.1993, in force 29.04.1997).
Convention With Respect to the Laws and Customs of War on Land (Hague Convention II), (enacted 29.07.1899, in force 04.09.1900).
Corn, G., 'Mixing Apples and Hand Grenades: The Logical Limit of Applying Human Rights Norms to Armed Conflict', *Journal of International Humanitarian Legal Studies*, no. 1, 2010, pp. 52–94.
Crawford, J. and J. Watkins, 'International Responsibility' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 283–95.

Declaration (IV,1), to Prohibit, for the Term of Five Years, the Launching of Projectiles and Explosives from Balloons, and Other Methods of Similar Nature. The Hague, 29.07.1899.
Dennett, D.C., 'When Hal Kills, Who's to Blame? Computer Ethics', in D.G. Stork (ed), *HAL's Legacy: 2001's Computer as Dream and Reality*, Cambridge, MA, MIT Press, 1998, pp. 351–65.
Dennett, D.C., *Brainchildren: Essays on Designing Minds*, Cambridge, MA, MIT Press, 1998.
Dill, J., 'Applying the Principle of Proportionality in Combat Operations', Policy Briefing, *Oxford Institute for Ethics, Law, and Armed Conflict*, 2010.
Dinstein, Y., *The Conduct of Hostilities Under the Law of International Armed Conflict*, Cambridge University Press, 2004.
Dinstein, Y., 'The Principle of Proportionality' in K.M. Larsen, C.G.G. Cooper and G. Nystuen (eds), *Searching for a "Principle of Humanity" in International Humanitarian Law*, Cambridge University Press, 2013, pp. 72–85.
Docherty, B., 'The Trouble With Killer Robots', *Foreign Policy*, 19.11.2012, http://www.foreignpolicy.com/articles/2012/11/19/the_trouble_with_killer_robots, (accessed 17.03.2014).
Doswald-Beck, L., 'International Humanitarian Law and the Advisory Opinion of the International Court of Justice on the Legality of the Threat or Use of Nuclear Weapons', International Committee of the Red Cross (ICRC), article no. 316, http://www.icrc.org/eng/resources/documents/misc/57jnfm.htm, 1997, (accessed 24.01.2014).
Dullum, O., 'Cluster Weapons; Military Utility and Alternatives', *Norwegian Defence Research Establishment* (FFI), report 02345, 2007.
Ekelhof, M. and M. Struyk, 'Deadly Decisions: 8 Objections to Killer Robots', *Pax for Peace*, 2014, http://www.paxforpeace.nl/media/files/deadlydecisionsweb.pdf, (accessed 14.05.2014).
Emmerson, B., 'Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism', Report of the Special Rapporteur, third Annual Report, United Nations General Assembly, A/68/389, 18.09.2013.
Engels, J. and W.O. Saas, 'On Acquiescence and End-Less War: An Inquiry into the New War Rhetoric', *Quarterly Journal of Speech*, vol. 99, no. 2, 2013, pp. 225–32.
Erasmus, *Collected Works of Erasmus: Adages*, trans. D.L. Drysdall, University of Toronto Press, 1982.
European Parliament Resolution on the Use of Armed Drones, Resolution 2567, 25.02.2014.
Farrelly, C., 'Justice in Ideal Theory: A Refutation', *Political Studies*, vol. 55, no. 4, 2007, pp. 844-64.
Fleck, D., 'The Law of Non-International Armed Conflicts' in D. Fleck (ed), *The Handbook of International Humanitarian Law,* 2nd edn, Oxford University Press, 2008, pp. 605–33.
Gadamer, H.G., *Truth and Method*, trans. J. Weinsheimer and D.G. Marshall, London, Sheed & Ward, 1999.
Galliott, J.C., 'Uninhabited Aerial Vehicles and the Asymmetry Objection: a Response to Strawser', *Journal of Military Ethics*, vol. 11, no. 1, 2012a, pp. 58–66.
Galliott, J.C., 'Viewpoint Article Closing With Completeness: The Asymmetric Drone Warfare Debate', *Journal of Military Ethics*, vol. 11, no. 4, 2012b, pp. 353–356.
Gardner, J., 'Legal Positivism: 5 ½ Myths', *American Journal of Jurisprudence*, vol. 46, no. 1, 2001, pp. 199–227.
Gasser, H.P., 'Protection of the Civilian Population' in D. Fleck (ed)*, The Handbook of International Humanitarian* Law, 2nd edn, Oxford University Press, 2008, pp. 237–319.
Gayle, B., 'US Navy "Stealth Drone" Takes to the Sea for Tests: The Autonomous X-47B Is Hoped to Be First Carrier-Borne Unmanned Aircraft', *Daily Mail*, 29.11.2012,

http://www.dailymail.co.uk/sciencetech/article-2240394/X-47B-stealth-drone-hoped-carrier-borne-unmanned-aircraft.html, (accessed 31.01.2014).
Geddes, B., 'How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics', *Political Analysis*, vol. 2, no. 1, 1990, pp. 131–50.
Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (enacted 12.08.1949, in force 21.10.1950).
Geneva Convention (III) relative to the Treatment of Prisoners of War (enacted 12.08.1949, in force 21.10.1950).
Gertler, J., 'US Unmanned Aerial Systems', *Congressional Research Service*, 7-5700, 03.01.2012.
Goodman, R., 'European Parliament Passes Resolution on Drones, Targeted Killings, and Fully Autonomous Weapons', *Just Security*, 27.02.2014, http://justsecurity.org/2014/02/27/european-parliament-passes-resolution-drones-targeted-killings-fully-autonomous-weapons/, (accessed 14.03.2014).
Goose, S., 'The Future of Global Warfare: Killer Robots', *Human Rights Watch Online*, 20.11.2012, http://www.hrw.org/news/2012/11/20/future-global-warfare-killer-robots, (accessed 07.08.2013).
Greenfield, D., 'The Case Against Drone Strikes Against People Who Only "Act" Like Terrorists', *The Atlantic*, 19.08.2013.
Greenwald, G., 'US Drone Strikes Target Rescuers in Pakistan – and the West Stays Silent', *The Guardian*, 20.08.2012.
Greenwood, C., 'Historical Development and Legal Basis' in D. Fleck (ed)*, The Handbook of International Humanitarian Law*, 2nd edn, Oxford University Press, 2008, pp. 1–44.
Gregory, D., 'From a View to Kill: Drones and Late Modern War', *Theory, Culture & Society*, vol. 28, no. 2, 2012, pp. 188–215.
Gye, H., '"I Count the bodies and watch the funerals... it's not just like PlayStation": Disturbing look inside the mind of a drone bomber pilot', *Daily Mail*, 09.06.2012.
Haghshenass, F., 'Iran's Asymmetric Naval Warfare', *The Washington Institute for Near East Policy*, Policy Focus 87, 2008, http://www.washingtoninstitute.org/uploads/Documents/pubs/PolicyFocus87.pdf, (accessed 27.11.2013).
Hamlin, A. and Z. Stemplowska, 'Theory, Ideal Theory, and the Theory of Ideals', *Political Studies Review*, vol. 10, no. 1, 2012, pp. 48–62.
Hammer, D., 'The Iliad as Ethical Thinking: Politics, Pity, and the Operation of Esteem', *Arethusa*, vol. 35, no. 2, 2002, pp. 203–35.
Hammes, T. X., *The Sling and the Stone: On Warfare in the 21st Century*, St. Paul, MN, Zenith Press, 2004.
Hammes, T.X., lecture at the Norwegian Institute of International Affairs (NUPI), 14.01.2014.
Hart, H.L.A., *The Concept of Law,* Oxford University Press, 1961.
Hegel, G.W.F., *Grundlinien der Philosophie des Rechts*, G.J.P.J. Bolland (ed), Leiden, A.H. Adriani [facsimile edition by Nabu Public Domain Reprints, s.a.], 1902 [1820].
Hegel, G.W.F., *Phänomenologie des Geistes*, G. Lasson (ed), Hamburg, Meiner, 1952 [1807].
Hegel, G.W.F., *Wissenschaft der Logik I*, G. Lasson (ed), Hamburg, Meiner, 1963 [1812–16].
Hegel, G.W.F., J*enaer Realphilosophie. Vorlesungsskripte zur Philosophie der Natur und des Geists von 1805–1806*, Johannes Hoffmeister (ed), Hamburg, Meiner, 1969 [1931].
Hegel, G.W.F., *Phenomenology of Spirit*, trans. A.V. Miller, Oxford University Press, 1977 [1807].
Hellström, T., 'On the Moral Responsibility of Military Robots', *Ethics of Information Technology*, vol. 14, no. 2, 2013, pp. 99–107.
Henckaerts, J.M., 'Study on Customary International Humanitarian Law: A Contribution to the Understanding and Respect for the Rule of Law in Armed Conflict', *International Review of the Red Cross*, vol. 87, no. 857, 2005, pp. 175–212.
Hendler, J., 'Computing Experts from 37 Countries Call Out for a Ban on Killer Robots',

International Campaign for Robot Arms Control (ICRAC), 16.10.2013, http://icrac.net/2013/10/computing-experts-from-37-countries-call-for-ban-on-killer-robots/, (accessed 14.03.2014).
Herodotus, *The History*, trans. G. Rawlinson, *Columbia University Libraries* (CLIO), 1860 [ca 420 BC].
Heyns, C., 'Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, United Nations Human Rights Council (UNHCR), Agenda Item 3, General Assembly, 09.04.2013.
Hitchens, C., 'American Inaction Favors Qaddafi', *Slate*, 07.03.2011.
Hobbes, T., *Leviathan*, R. Tuck (ed), Cambridge University Press, 2008 [1651].
Homer, *The Iliad*, trans. E.V. Rieu (revised and updated with P. Jones), London, Penguin, 2003a.
Homer, *The Odyssey*, trans. E.V. Rieu, London, Penguin, 2003b.
Honneth, A., 'Integrity and Disrespect: Principles of a Conception of Morality Based on the Theory of Recognition', *Political Theory*, vol. 20, no. 2, 1992, pp. 187–201.
Honneth, A., *The Struggle for Recognition: The Moral Grammar of Social Conflict*, trans. J. Anderson, Cambridge, MA, MIT Press, 1995 [1992].
Honneth, A., 'Umverteilung als Anerkennung: Eine Erwiderung auf Nancy Fraser' in N. Fraser and Axel Honneth, *Umverteilung oder Anerkennung? Eine politisch-philosophische Kontroverse*, Frankfurt am Main, Suhrkamp, 2003.
Honneth, A., *Das Recht der Freiheit*, Frankfurt am Main, Suhrkamp, 2011.
Horkheimer, M., *Traditionelle und Kritische Theorie*, Frankfurt am Main, Fischer Bücherei, 1970 [1937], pp. 1–64.
Hughes, S., 'Campaigners Call for an International Ban on "Killer Robots"', *BBC News*, 23.04.2013, http://www.bbc.co.uk/news/uk-22250664, (accessed 31.01.2014).
Human Rights Watch, 'Losing Humanity: The Case Against Killer Robots', International Human Rights Clinic (IHRC), Harvard Law School, 2012.
Hume, D., *An Enquiry Concerning Human Understan*ding, Mineola, NY, Dover, 2004.
ICRC, (International Committee of the Red Cross), 'Anti-personnel Mines: Friend or Foe? A Study of the Military Use and Effectiveness of Anti-Personnel Mines', Legal Division, 1994.
ICRC (International Committee of the Red Cross), 'Fully Autonomous Weapon System', 25.11.2013, http://www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm, (accessed 30.01.2014).
ICRC (International Committee of the Red Cross), 'Customary International Law: Practice Relating to Rule 14 Proportionality in Attack', 2014a, http://www.icrc.org/customary-ihl/eng/docs/v2_cha_chapter4_rule14, (accessed 26.03.2014).
ICRC (International Committee of the Red Cross), 'Report of the ICRC Expert Meeting on 'Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects', 26–28 March 2014, Geneva, 2014b. Document available upon request.
Ignatieff, M., 'Drones Give Democracies No Cause for War', *Financial Times*, 12.06.2012, http://www.ft.com/intl/cms/s/0/10a03278-b3b3-11e1-a3db-00144feabdc0.html#axzz2xAtowHYH, (accessed 27.03.2014).
International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law, 'Living Under Drones: Death, Injury, and Trauma to Civilians From US Drone Practices in Pakistan', 2012.
Ipsen, K., 'Combatants and Non-Combatants' in D. Fleck (ed)*, The Handbook of International Humanitarian Law*, 2nd edn, Oxford University Press, 2008, pp. 79–117.
Iqbal, A., 'Few Things are More Inherently Un-American Than Drones', *Dawn News*, 17.11.2012, http://beta.dawn.com/news/764782/few-things-are-more-inherently-un-american-than-drones, (accessed 20.08.2013).

Japanese Ministry of Foreign Affairs, 'Statement by H.E. Ambassador Toshio Sano, Permanent representative to the Conference on Disarmament ', Experts Meeting on Lethal Autonomous Weapon Systems to the CCW, 13.05.2014.
Johnson, A.M. and S. Axinn, 'The Morality of Autonomous Robots', *Journal of Military Ethics*, vol. 12, no. 2, 2013, pp. 129–41.
Johnson, R., 'Defending Ways of Life: The (Anti-)Terrorist Rhetorics of Bush and Blair', *Theory, Culture & Society*, vol. 19, no. 4, 2002, pp. 211–31.
Joint Service Manual of the Law of Armed Conflict (UK Manual), Joint Service publication 383, 2014.
Kadish, S., *Blame and Punishment: Essay's in the Criminal Law*, London, Macmillan, 1987.
Kahn, P.W., 'The Paradox of Riskless Warfare', *Philosophy and Public Policy Quarterly* vol. 22, no 3, 2002, pp. 2–8.
Kant, I., *Grundlegung zur Metaphysik der Sitten*, Cologne, Anaconda, 2008 [1785].
Kels, C.G., 'Closing Argument: At the Outer Bounds of Asymmetry', *Journal of Military Ethics*, vol. 11, no. 3, 2012, pp. 223–44.
Kilcullen, D. and A. McDonald Exum, 'Death From Above, Outrage Down Below', *The New York Times*, 16.05.2009, http://www.nytimes.com/2009/05/17/opinion/17exum.html?pagewanted=all&_r=0, (accessed 03.10.2013).
King, G., R.O. Keohane and S. Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton University Press, 1994.
Kojève, A., *Introduction to the Reading of Hegel*, trans. J.H. Nichols jr., London, Cornell University Press, 1980.
Krigskorset med sverd, statutter, innstiftet 23. mai 1941, Direktiv for bruk of bæring av dekorasjoner, ordener og medaljer i Forsvaret, 26.04.2012.
Krishnan, A., *Killer Robots: Legality and Ethicality of Autonomous Weapons*, Farnham, Ashgate, 2009.
Lakatos, I., 'Falsification and the Methodology of Scientific Research Programmes' in I.
Lakatos and A. Musgrave (eds). *Criticism and the Growth of Knowledge*, Cambridge University Press, 1984 [1970], pp. 91–196.
Larsen, K.M., C.G.G. Cooper and G. Nystuen, 'Introduction' in K.M. Larsen, C.G.G. Cooper and G. Nystuen (eds), *Searching for a "Principle of Humanity" in International Humanitarian Law*, Cambridge University Press, 2013, pp. 1–19.
Legality of the Threat or Use of Nuclear Weapons, International Court of Justice Advisory Opinion, 08.07.1996.
Leopold, D., 'Dialectical Approaches', in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches*, Oxford University Press, 2010, pp. 106–27.
Lin, P., G. Bekey and K. Abney, 'Autonomous Military Robotics: Risk, Ethics, and Design'. Report Prepared for the US Department of Navy*, Office of Naval Research*. California Polytechnic State University, 2008, www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA534697, (accessed 08.09.2013).
Liu, H.Y., 'Categorization and Legality of Autonomous and Remote Weapons systems', *International Review of the Red Cross*, vol. 94, no. 886, 2012, pp. 627–52.
Luban, D., 'International Criminal Law' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 569–88.
Martinez, J.S. 'Understanding *Mens Rea* in Command Responsibility: From *Yamashita* to *Blaškić* and Beyond', *Journal of International Criminal Justice*, vol. 5, no. 3, 2007, pp. 638–64.
Matthews, H., 'The Interaction Between International Human Rights Law and International Humanitarian Law: Seeking the Most Effective Protection of Civilians in Non-International Armed Conflicts', *International Journal of Human Rights*, vol. 17, no. 5–6, 2013, pp. 633–45.
Marchuk, I., *The Fundamental Concept of Crime in International Criminal Law*, Berlin, Springer, 2014.
McMahan, J., *Killing in War*, Oxford University Press, 2009.

McMahan, J., 'Laws of War' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 493–509.
McNay, L., *Against Recognition*, Cambridge, Polity, 2008.
McNay, L., 'Recognition as Fact and Norm: the Method of Critique' in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches*, Oxford University Press, 2010, pp. 85–105.
McNeal, G., 'Fear and the Killer Robots Meme', *Forbes*, 27.02.2013, http://www.forbes.com/sites/gregorymcneal/2013/02/27/fear-and-the-killer-robots-meme/, (accessed 31.01.2014).
Mearsheimer, J.J., *The Tragedy of Great Power Politics*, New York, NY, W.W.Norton, 2001.
Meckstroth, C., 'Socratic Method and Political Science', *American Political Science Review*, vol. 106, no. 3, 2012, pp. 644–60.
Mick, J., 'Iran: Yes, We Hacked US's Drone, and Here's How We Did It', *Dailytech*, 15.12.2011, http://www.dailytech.com/Iran+Yes+We+Hacked+the+USs+Drone+and+Heres+How+We+Did+It/article23533.htm, (accessed 26.03.2014).
Mines Action Canada, *Banning Cluster Munitions: Government Policy and Practice*, Toronto, Joseph Communications.
Miller, G., 'Lawmakers Seek to Stymie Plan to Shift Control Drone Campaign from CIA to Pentagon', *Washington Post*, 16.01.2014.
Monbiot, G., 'With Its Deadly Drones, the US is Fighting a Coward's War', *The Guardian*, 30.01.2012, http://www.theguardian.com/commentisfree/2012/jan/30/deadly-drones-us-cowards-war. (accessed 10.08.2013).
Morkevicius, V., 'Tin Men: Ethics, Cybernetics and the Importance of Soul', *Journal of Military Ethics*, vol. 13, no. 1, 2014, pp. 3–19.
Montgomery, L., 'Younger Military Veterans Are Angered by Budget Cuts to Their Pension Benefits', *Washington Post*, 31.12.2013.
Morgenthau, H., *Politics Among Nations: The Struggle for Power and Peace*, New York, NY, McGraw Hill, 2006 [1948].
Mueller, G.E., 'The Hegel Legend of "Thesis–Antithesis–Synthesis"', *Journal of the History of Ideas*, vol. 19, no. 3, 1958, pp. 411–14.
Mulholland, J.A., *Kant's System of Rights*, New York, Columbia University Press, 1990.
Nagel, T., 'War and Massacre', *Philosophy and Public Affairs*, vol. 1, no 2, 1972, pp. 123–44.
Nahem, J., and K. Egeland, 'Legalitet og legitimitet: *responsibility to protect* i teori og praksis, *Internasjonal Politikk*, vol. 72, no. 1, 2014, pp. 102–12.
Naval Air Systems Command, 'X-47B Aircraft Recognized with Popular Mechanic Breakthrough Innovator Award', 11.11.2013, http://www.navair.navy.mil/index.cfm?fuseaction=home.NAVAIRNewsStory&id=5470, (accessed 31.01.2014).
Norpoth, H. and A.H. Sidman, 'Mission Accomplished: The Wartime Election in 2004', *Political Behaviour*, vol. 29, no. 2, 2007, pp. 175–95.
Norwegian Ministry of Foreign Affairs, 'General Statement by Norway', CCW Meeting of Experts on Lethal Autonomous Weapon Systems, 13–15 May 2014.
Nystuen, G. and K. Egeland, 'The Potential of the Arms Trade Treaty to Reduce Violations of International Humanitarian Law and Human Rights Law' in C.M. Bailliet and K.M. Larsen (eds), *Promoting Peace Through International Law*, Oxford University Press, forthcoming 2014.
O'Connell, M.E., 'Unlawful Killing With Combat Drones: A Case Study of Pakistan, 2004 2009' in S. Bronitt, M. Gani, and S. Hufnagel (eds), *Shooting to Kill: Socio-Legal Perspectives on the Use of Lethal Force*, Oxford, Hart, 2012, pp.263–91.
Oeter, S., 'Methods and Means in Combat' in D. Fleck (ed), *The Handbook of International Humanitarian Law*, 2nd edn, Oxford University Press, 2008, pp. 401–79.
O'Meara, R.M., 'Contemporary Governance Architecture Regarding Robotics Technologies: An Assessment' in P. Lin, K. Abney, and G.A. Bekey (eds), *Robot Ethics: The*

Ethical and Social Implications of Robotics, Cambridge, MA, MIT Press, 2012, pp. 159–168.
Orwell, G., 'Shooting an Elephant' in *A Collection of Essays*, London, Harvest, 1981 [1935].
Orwell, G., 'You and the Atomic Bomb' in Soina Orwell and Ian Angus (eds), *In Front of Your Nose, 1945–50: The Collected Essays, Journals and Letters of George Orwell, Vol. IV*, London, Secker & Warburg, 1968 [1945], pp. 6–10.
Otto, J.L. and B.J. Webber, 'Mental Health Diagnoses and Counselling Among Pilots of Remotely Piloted Aircraft in the United States Air Force', *Medical Surveillance Monthly Report (MSMR): A Publication of the Armed Forces Health Surveillance Center*', vol. 20, no. 3, 2013, pp. 3–8.
Owen, J., 'Anger at UK's Investment in "Risk-Free" Drone War', *The Independent*, 28.04.2013.
OED (*Oxford English Dictionary Online*), entries:
'Algorithm',
http://www.oxforddictionaries.com/definition/english/algorithm?q=algorithm, (accessed 05.04.2014).
'Android',
http://www.oxforddictionaries.com/definition/english/android?q=android, (accessed 05.04.2014).
Parks, W.H., 'Command Responsibility for War Crimes', *Military Law Review*, vol. 63, 1973, pp. 1–104, p. 37.
Philip, M., 'Political Theory and History' in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches*, Oxford University Press, 2010, pp. 128–49.
Pincus, W., 'The Inevitable Blowback to High-Tech Warfare', *The Washington Post*, 16.10.2012.
Plato, 'The Republic', trans. P. Shorey in E. Hamilton and H. Cairns (eds), *The Collected Dialogues of Plato*, Princeton University Press, 1989, pp. 575–844.
Popper, K., *The Open Society and Its Enemies. The High Tide of Prophecy: Hegel, Marx, and the Aftermath*, London, Routledge, 1945.
Primoratz, I., 'Walzer's Just War Theory: Some Issues of Responsibility', *Ethical Theory and Moral Practice*, vol. 5, no. 2, 2002, pp. 221–43.
Prosecutor v. Delalić et al. (Čelebići), Judgment, Case No. IT-96-21-T, 21.11.1998.
Prosecutor v. Ignace Bagilishema (Appeal Judgement: Reasons), International Criminal Tribunal for Rwanda (ICTR), 03.07.2002.
Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), (enacted 08.06.1977, in force 07.12.1977).
Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II), 8 June 1977 (enacted 08.06.1977, in force 07.12.1978).
Quince, A., 'Future Drone Strikes Could See Execution by Algorithm', *Australian Broadcasting Corporation* (ABC) online, 21.01.2013, http://www.abc.net.au/radionational/programs/rearvision/drones/4703792, (accessed 29.01.2014).
Rawls, J., *A Theory of Justice*, revised edn, Cambridge, MA, Harvard University Press, 1999 [1971].
Raytheon, 'Phalanx Close-In Weapon System', 2014, http://www.raytheon.com/capabilities/products/phalanx/, (accessed 28.01.2014).
Reaching Critical Will, 'Fully Autonomous Weapons', 2013, http://www.reachingcriticalwill.org/resources/fact-sheets/critical-issues/7972-fully-autonomous-weapons, (accessed 09.01.2014).
Rehman, J., *International Human Rights Law*, 2nd edn, Harlow, Pearson, 2010.
Reprieve, 'Drones', 2014, http://www.reprieve.org.uk/investigations/drones/, (accessed 14.03.2014).

Resolution RC/Res.6, 'Amendments to the Rome Statute of the International Criminal Court on the Crime of Aggression' 11.06.2010.
Rome Statute of the International Criminal Court (ICC), (enacted 17.07.1998, in force 01.07.2002).
Rose, P., 'Divinising Technology and Violence: Technopoly, the Warfare State, and the Revolution in Military Affairs, *Journal of Contemporary Religion*, vol. 27, no. 3, 2012, pp. 265–81.
Ruud, M. and G. Ulfstein, *Innføring i folkerett*, 3rd edn, Oslo, Universitetsforlaget, 2006.
Ruge, C.H., 'Innledning' in C.H. Ruge (ed), *En framtid uten miner: 10 år med minekonvensjonen*, Oslo, Conflux, 2007, pp. 15–26.
Salisbury, J.E., *The Beast Within: Animals in the Middle Ages*, 2nd edn, New York, Routledge, 2011.
Sartre, J.P., *Being and Nothingness*, trans. H.E. Barnes, London, Routledge, 2010 [1943].
Sassòli, M., 'Autonomous Weapons – Potential Advantages for the Respect of International Humanitarian law', *Professionals in Humanitarian Assistance and Protection*, 2013, http://phap.org/articles/autonomous-weapons-–-potential-advantages-respect-international-humanitarian-law, (accessed 26.03.2014).
Scahill, J., 'The Secret US War in Pakistan', *The Nation*, 23.11.2009, http://www.thenation.com/article/secret-us-war-pakistan, (accessed 27.03.2014).
Scahill, J. and G. Greenwald, 'The NSA's Secret Role in the U.S. Assassination Program', *The Intercept*, 10.02.2014, https://firstlook.org/theintercept/article/2014/02/10/the-nsas-secret-role/, (accessed 14.02.2014).
Schmitt, M.N., 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics', *National Security Journal*, online feature, http://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/, 2013, (accessed 09.12.2013).
Schmitt, M.N., 'Military Necessity and Humanity in International Humanitarian Law: Preserving the Delicate Balance', *Virginia Journal of International Law*, vol. 50, no. 4, 2010, pp. 795–839.
Schmitt, M.N. and J.S. Thurnher, '"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict', *Harvard National Security Journal*, vol. 4, no. 2, 2013, pp. 231–81.
Shah, M.A., 'Activists Campaign in European Cities Against Drone Strikes', *The News*, 22.02.2014, http://www.thenews.com.pk/Todays-News-2-234193-Activists-campaign-in-European-cities-against-drone-strikes, (accessed 14.03.2014).
Shakespeare, W., *Romeo and Juliet*, Hertfordshire, Wordsworth, 2000 [1597].
Shane, S., 'U.S. Said to Target Rescuers at Drone Strike Sites', *New York Times*, 05.02.2012.
Sharkey, N., 'Automated Killers and the Computing Profession', *Computer*, vol. 40, no. 11, 2007, pp. 122–24.
Sharkey, N., 'Cassandra or False Prophet of Doom: AI Robots and War', *IEEE Intelligent Systems*, 2008, pp. 14–17.
Sharkey, N., 'Saying "No!" to Lethal Autonomous Targeting', *Journal of Military Ethics*, vol. 9, no. 4, 2010, pp. 369–83.
Sharkey, N., 'Noel Sharkey on Drones and the Threat of Autonomous Weapons', *New World*, 19.04.2013, http://www.una.org.uk/magazine/spring-2013/noel-sharkey-drones-and-threat-autonomous-weapons, (accessed 31.01.2014).
Shell, S.M., *Kant and the Limits of Autonomy*, Cambridge, MA, Harvard University Press, 2009.
Shue, H., 'Laws of War' in S. Besson and J. Tasioulas (eds), *The Philosophy of International Law*, Oxford University Press, 2010, pp. 511–27.
Singer, P.W., *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, New York, Penguin, 2010.
Sliedregt, E.V., 'Command Responsibility at the ICTY – Three Generations and Still Ambiguity' in A.H. Swart et al., *The Legacy of the ICTY*, Oxford University Press, 2011, pp. 377–400.

Snyder, G.H., 'Mearsheimer's World – Offensive Realism and the Struggle for Security', *International Security*, vol. 27, no. 1, 2002, pp. 149–73.
Sokolowski, R., 'Friendship and Moral Action in Aristotle', *The Journal of Value Inquiry*, vol. 35, no. 3, 2001, pp. 355–69.
Sparrow, R., 'Killer Robots', *Journal of Applied Philosophy*, vol. 24, no. 1, 2007, pp. 62–77.
Speight, A., 'Arendt and Hegel on the Tragic Nature of Action', *Philosophy & Social Criticism*, vol. 27, no. 5, 2002, pp. 523–36.
Spenser, T., 'New Biometric Gait Recognition Research Conducted in U.K.', *Biometric*, 25.09.2012.
Stanford and NYU, 'Living Under Drones: Death, Injury, and Trauma to Civilians from US Drone Practices in Pakistan', International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law, 2012.
Statute for the International Court of Justice (ICJ), annexed to the Charter of the United Nations (enacted 26.06.1945, in force 24.10.1945).
Stemplowska, S., 'What's Ideal About Ideal Theory?', *Social Theory and Practice*, vol. 34, no. 3, 2008, pp. 319-40.
Stephan, K.D., K. Michael, M.G. Michael, L. Jacob, and E.P. Anesta, 'Social Implications of Technology: The Past, the Present, and the Future', *Proceedings of the IEEE*, vol. 100, 2012, pp. 1752–81.
Stop Killer Robots, official website, http://www.stopkillerrobots.org/.
Strawser, B.J., 'Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles', *Journal of Military Ethics*, vol. 9, no. 4, 2010, pp. 342–68.
Sturgeon, N.L., 'Hume's Metaethics' in R.E. Radcliffe (ed), *A Companion to Hume*, Chichester, Wiley-Blackwell, 2011.
Sullins, J.P., 'When is a Robot a Moral Agent?', *International Review of Information Ethics*, vol. 6, December 2006, pp. 23–30.
Swift, A. and S. White, 'Political Theory, Social Science, and Real Politics' in D. Leopold and M. Stears (eds), *Political Theory: Methods and Approaches*, Oxford University Press, 2010, pp. 48–69.
Swiss Ministry of Foreign Affairs, 'Réunion d'Experts de la Convention sur Certaines Armes Classiques', Geneva 13.05.2014.
Syse, H., *Rettferdig Krig?,* Oslo, Aschehoug, 2003.
Taylor, C., 'The Politics of Recognition' in A. Gutmann (ed), *Multiculturalism: Examining the Politics of Recognition*, Princeton, NJ, Princeton University Press, 1994, pp. 25–73.
Taylor, C., *Hegel*, Cambridge University Press, 2005 [1975].
Thies, W.J., *Why NATO Endures*, Cambridge University Press, 2009.
Thucydides, *History of the Peloponnesian War*, trans. R. Crawley, Project Gutenberg, 2009 [410 BC].
Tomuschat, C., 'Human Rights and International Humanitarian Law', *European Journal of International Law*, vol. 21, no. 1, 2010, pp. 15–23.
Tonkens, R., 'The Case Against Robotic Warfare: a Response to Arkin', *Journal of Military Ethics*, vol. 11, no. 2, pp. 149–68.
United Nations General Assembly First Committee official website, '"Record-Breaking" Participation Defined First Committee Session, Says Chair at Closing, as Delegates Send 53 Drafts to General Assembly', 5 November 2013, http://www.un.org/News/Press/docs/2013/gadis3494.doc.htm, (accessed 07.01.2014).
United Nations News Centre, 'UN Human Rights Expert Urges Global Pause in Creation of Robots With "Power to Kill"', http://www.un.org/apps/news/story.asp?NewsID=45042#.Usu1FrRWPq4, (accessed 07.01.2014).
United Nations Office for Disarmament Affairs (UNODA) official website, 'High Representative's Statement to the Human Rights Council on the Topic of Lethal

Autonomous Robotics', http://www.un.org/disarmament/update/20130530/, (accessed 07.01.2014).
Universal Declaration of Human Rights (UDHR), adopted by UN General Assembly resolution 217 (III), (enacted 10.12.1948).
US Department of Defense, Department of Defense Directive (DoDD) number 3000.09, November 21, 2012.
US Department of Defense, 'Unmanned Systems Integrated Roadmap', 2013, http://www.defense.gov/pubs/DOD-USRM-2013.pdf, (accessed 10.05.2014).
US Joint Forces Command (USJFCOM), 'Joint Innovation and Experimentation Directorate Briefing to Precision Strike Association', 2007, http://www.dtic.mil/ndia/2007psa_winter/winnefeld.pdf, (accessed 28.02.2014).
Van Evera, S., *Causes of War: Power and the Roots of Conflict*, Ithaca, NY, Cornell University Press, 1999.
Vetlesen, A.J. 'Innledning' in H. Arendt, *Vita Activa*, Oslo, Pax, 1996.
Vienna Convention on the Law of Treaties (VCLT), (enacted 23.05.1969, in force 27.01.1980).
Wallach, W. and C. Allen, 'Moral Machines: Contradiction in Terms or Abdication of Human Responsibility?' in P. Lin, K. Abney, and G. Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics*, Cambridge, MA, MIT Press, 2012, pp. 55–68.
Wallach, W. and C. Allen, 'Framing Robot Arms Control', *Ethics and Information Technology*, vol. 15, no. 2, 2013, pp. 125–35.
Walter, J. (ed), *Kindlers Literaturlexikon*, vol. 3, München, Kindler Verlag, 1989.
Walzer, M., *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 4th edn, New York, NY, Basic Books, 2006 [1977].
Walzer, M., *Interpretation and Social Criticism*, Cambridge, MA, Harvard University Press, 1987.
Weiner, T., 'A New Model Army Soldier Rolls Closer to the Battlefield', *The New York Times*, 16.02.2013.
Wheeler, W., 'The MQ-9's Cost and Performance', *Time*, 28.02.2012.
Whitlock, C., 'U.S. Airstrike that Killed American teen in Yemen Raises Legal, Ethical Questions', *Washington Post*, 22.10.2011.
Williams, R.R., *Hegel's Ethics of Recognition*, Berkely, University of California Press, 1997.
Williamson, J.A., 'Some Considerations on Command Responsibility and Criminal Liability', *International Review of the Red Cross*, vol. 90, no. 870, pp. 303–17.
Wolfendale, J., 'Professional Integrity and Disobedience in the Military', *Journal of Military Ethics,* vol. 8, no. 2, 2009, pp. 127–40.
Wolin, S.S., *Politics and Vision: Continuity and Innovation in Western Political Thought*, expanded edn, Princeton, NJ, Princeton University Press, 2004 [1960].
Wolin, S.S., 'Political Theory as a Vocation', *American Political Science Review*, vol. 63, no. 4, 1969, pp. 1062–82.
Yar, M., 'Hannah Arendt (1906–1975)', *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/arendt/#H6, 22.07.2005, (accessed 16.11.2013).
Zenko, M., 'Reforming US. Drone Strike Policies', *Council of Foreign Relations*, Special Report No. 65, 2013.